



Universidade do Minho
Escola de Engenharia

Cláudio Filipe Ferreira Nóvoa

**Melhorar a taxa de fermentação da xilose
em etanol em *Saccharomyces cerevisiae*
através da incorporação de informações
estruturais de proteínas**

Tese de Mestrado

Bioinformática

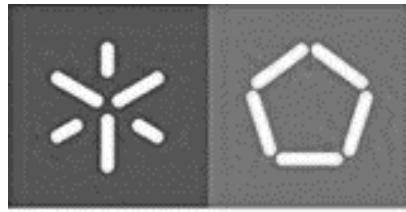
Trabalho efetuado sob a orientação do(s)

Doutora Isabel Rocha

Doutor Cláudio Soares

Janeiro

2019



Universidade do Minho
Escola de Engenharia

Melhorar a taxa de fermentação da xilose em etanol em *Saccharomyces cerevisiae* através da incorporação de informações estruturais de proteínas

Mestrado em Bioinformática

Trabalho realizado sob a orientação de
Professora Doutora Isabel Rocha
Professor Doutor Cláudio Soares

Janeiro de 2019



DECLARAÇÃO

Nome: Cláudio Filipe Ferreira Nóvoa

Endereço eletrónico: claudio_ffn@hotmail.com

Telefone: 963536652

Número do Bilhete de Identidade: 14913157

Título da dissertação: Melhorar a taxa de fermentação da D-xilose em etanol em *Saccharomyces cerevisiae* incorporando informações estruturais de proteínas

Orientador(es): Isabel Cristina Almeida Pereira Rocha; Cláudio M. Soares

Ano de conclusão: 2019

Designação do Mestrado: Bioinformática

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA DISSERTAÇÃO APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE;

Universidade do Minho, 31/01/2019

Assinatura:

Cláudio Filipe Ferreira Nóvoa



AGRADECIMENTOS

Em primeiro lugar agradeço à minha orientadora, a Professora Doutora Isabel Rocha por me ter proporcionado a oportunidade de realizar este trabalho, e também, ao mar de conhecimentos que adquiri com ela e ao tempo que dispensou para a realização deste trabalho. Agradeço também ao Professor Doutor Cláudio Soares por me ter proporcionado a oportunidade de envergar nesta área.

Ao Tiago Resende, deixo aqui um obrigado especial por todo o tempo, paciência, boa disposição, disponibilidade e tudo que me ensinou ao longo deste ano.

Agradeço também à Joana Oliveira por toda a ajuda e simpatia ao longo deste trabalho.

À minha família, principalmente aos meus pais, pela confiança que depositaram em mim e também pelo apoio que me deram das mais variadas formas.

Aos meus amigos, que duma forma indireta contribuíram para a realização deste trabalho, através do incentivo e companheirismo.

E por fim, à Escola de Engenharia e a todos os docentes do mestrado em bioinformática por toda a dedicação e esforço para ajudar a formar bioinformáticos competentes e capazes.





RESUMO

O consumo crescente do petróleo constitui um grave problema ambiental e económico para a sociedade atual. Para solucionar este problema, o uso de biocombustíveis apresenta-se como uma viável alternativa. Através do aproveitamento da cana-de-açúcar e do milho, utilizando microrganismos, é possível obter etanol, o biocombustível mais utilizado atualmente. Porém, nem todos os açúcares provenientes destas matérias-primas são fermentados de forma economicamente eficiente, sendo que a D-xilose, o segundo açúcar mais abundante, continua a ter um aproveitamento ineficiente por parte de microrganismos como a *Saccharomyce cerevisiae*, apesar das alternativas já exploradas. A via de fermentação da xilose é constituída por diversas enzimas, sendo que o problema da fraca taxa de conversão de xilose em etanol se deva essencialmente às primeiras duas: a xilose redutase e a xilitol desidrogenase.

Muitas das abordagens para solucionar o ineficiente consumo da D-xilose têm por base a engenharia metabólica, ou como tornar a xilose redutase ou a xilitol desidrogenase específica para o mesmo cofator. Este trabalho tem o objetivo de descobrir características estruturais, na xilose redutase e na xilitol desidrogenase, que possam ter influência na afinidade da D-xilose e no seu modo de ligação. Para tal foram usadas e testadas abordagens *in silico* tendo em consideração a estrutura das proteínas e os açúcares que estas utilizam como substrato, deixando para segundo plano os cofatores e o metabolismo. Foram recolhidas diversas sequências de xilose redutases e xilitol desidrogenases de diversos organismos e as suas estruturas tridimensionais modeladas. Para esses modelos foi medido o volume dos seus centros ativos e realizado o *docking* molecular da D-xilose, xilitol e outros substratos. Os resultados do *docking* e dos volumes foram comparados com os K_M dos diferentes substratos.

Praticamente todos os volumes da xilose redutase foram corretamente medidos, o que não se verificou no caso da xilitol desidrogenase. Relativamente ao *docking* molecular, foram analisados os resíduos envolvidos nos processos catalíticos destas duas enzimas bem como os scores resultantes. A *Candida tenuis* e *Candida boidiini* foram as que apresentaram uma maior afinidade no *docking* da xilose com um valor de 5,1. No *docking* do xilitol nas xilitol desidrogenases a *Rhizomucor pusillu* foi a que teve melhor score com 5,1.

Este trabalho evidenciou que os volumes do centro ativo e o K_M aparentam não estar diretamente relacionados. Os resíduos N e H do centro catalítico da xilose redutase estão



conservados em todas as xilose redutases e estão envolvidos nas interações polares com a xilose e outros açúcares.

PALAVRAS-CHAVE

Xilose redutase, Xilitol desidrogenase, D-xilose, xilitol, *docking*, volume



ABSTRACT

The growing consumption of oil is a serious environmental and economic problem for today's society. To solve this problem, the use of biofuels presents itself as a viable alternative. Through the use of sugar cane and corn, using microorganisms, it is possible to obtain ethanol, the most commonly used biofuel. However, not all sugars from these raw materials are fermented in an economically efficient manner, and D-xylose, the second most abundant sugar, continues to be inefficiently utilized by microorganisms such as *Saccharomyces cerevisiae*, despite of the alternatives already available and explored. The xylose fermentation path consists of several enzymes, and the problem of the low conversion rate of xylose into ethanol is due to the first two: xylose reductase and xylitol dehydrogenase.

The approaches to solving the inefficient consumption of D-xylose are mainly based on metabolic engineering, such as turning xylose reductase or xylitol dehydrogenase specific for the same cofactor. In this work, in order to discover structural features in xylose reductase and xylitol dehydrogenase that may influence the affinity of D-xylose and its mode of binding, *in silico* approaches were used and tested taking into account the structure of proteins and sugars which they use as a substrate, leaving the cofactors and the metabolism behind. Several xylose reductase and xylitol dehydrogenase sequences from various organisms were collected and their tridimensional structure modelled. For these models the volume of its active centers was measured and the molecular docking of D-xylose, xylitol and other substrates was performed. The results of the docking and the volumes were compared with the K_M of the different substrates.

Almost all volumes of xylose reductase were correctly measured, which was not observed in the case of xylitol dehydrogenase. Regarding molecular docking, the residues involved in the catalytic processes of these two enzymes as well as the resulting scores were analyzed. *Candida tenuis* and *Candida boidini* had the highest affinity for xylose docking with a value of 5.1. In the docking of xylitol in xylitol dehydrogenases *Rhizomucor pusillus* was the one that had the best score with.

This work evidenced that the volumes of the active center and the K_M are not directly related. The N and H residues of the catalytic center of xylose reductase are conserved in all xylose reductases and are involved in polar interactions with xylose and other sugars.



KEYWORDS

Xylose reductase, xylitol dehydrogenase, D-xylose, xylitol, volume



ÍNDICE

Agradecimentos.....	iii
Resumo	v
Abstract.....	vii
Índice de Figuras	xi
Índice de Tabelas.....	xv
Lista de Abreviaturas, Siglas e Acrónimos.....	xvii
1. Introdução/Estado da arte	1
1.1 Modificação genética de <i>Saccharomyces cerevisiae</i>	3
1.2 Problemas com a <i>Saccharomyces cerevisiae</i> modificada com a via XI e XR/XDH e soluções já exploradas.....	5
1.2.1 Modificação dos genes que codificam os transportadores de hexoses para melhorar a entrada e consumo da D-xilose.....	5
1.2.2 A acumulação de xilitol e trocas de afinidades dos cofatores	6
1.3 Biologia estrutural	7
1.3.1 Centro ativo de Enzimas.....	9
1.3.2 Métodos computacionais para identificação de centros ativos e análise das suas propriedades	10
1.3.3 Molecular Dynamics e MD simulations.....	13
1.3.4 Molecular Docking.....	13
1.3.5 As duas enzimas principais da via de fermentação da xilose	16
1.3.6 Análises estruturais às enzimas da via XR/XDH.....	18
2. Materiais e métodos.....	21
2.1 Recolha de dados e informação.....	21
2.2 Construção dos modelos das enzimas	22
2.3 Medição dos volumes dos centros ativos das enzimas	23
2.4 Docking.....	23
2.5 Visualização dos aminoácidos e interações no centro ativo	24
3. Resultados e discussão.....	25
3.1 Dados da xilose redutase (XR).....	25
3.2 Estruturas tridimensionais das xilose redutases	29



3.3	Dados da xilitol desidrogenase	31
3.4	Estruturas tridimensionais das xilitol desidrogenases	33
3.5	Volumes dos centros ativos das xilose redutases e xilitol desidrogenases.....	34
3.6	Volume dos vários mutantes da levedura <i>Scheffersomyces stipitis</i>	37
3.7	Volumes dos centros ativos das xilitol desidrogenases	39
3.8	Docking da xilose nas xilose redutases	40
3.9	A influência do volume do centro ativo nos resultados do <i>docking</i>	41
3.10	Docking molecular na enzima XR da <i>Saccharomyces cerevisiae</i>	41
3.10.1	Visualização dos resultados do docking da <i>Saccharomyces cerevisiae</i>	43
3.11	Docking do xilitol nas xilitol desidrogenases: propriedades, análise visual e química	44
4.	Conclusão e perspectivas futuras	48
REFERÊNCIAS BIBLIOGRÁFICAS		49



ÍNDICE DE FIGURAS

Figura 1 - Molécula de etanol e os seus respetivos grupos funcionais: colorido a vermelho o grupo metil (CH ₃), um grupo metileno (CH ₂) a azul e um grupo hidroxilo (OH) a preto [5]. ..	2
Figura 2 - Estrutura de alguns dos açúcares obtidos a partir da hidrólise da biomassa lignocelulósica e algumas das suas possíveis reações [7].	2
Figura 3 - A via de fermentação de xilose na <i>Saccharomyces cerevisiae</i> recombinante [3].....	4
Figura 4 - Esquema do processo de modelação de proteínas por homologia [50].....	9
Figura 5 - Metodologia de procura do algoritmo usado pelo trj_cavity (A) e representação de um voxel e as suas possíveis direções de procura (B) [52].	12
Figura 6 - Exemplo das diferentes conformações da mesma família de proteínas sobrepostas mas com diferentes substratos ligados [81].	13
Figura 7 - Mecanismo de procura dos GAs [99].....	16
Figura 8 - Disposição espacial dos resíduos Asp-50, Tyr-55, Lys-84 e His-117 presentes no centro catalítico das AKR (subfamília da Xilose Reductase) e as suas interações com a molécula de água (preto) e o anel de nicotinamida do NAD (rosa) [106].....	17
Figura 9 - Mecanismos de reação da xilose redutase da <i>Candida tenuis</i> e os resíduos envolvidos na reação [112].	19
Figura 10 - Esquema e sequência temporal das metodologias usadas para a elaboração deste trabalho.	21
Figura 11- Gráfico com os diferentes organismos e o número respetivo de sequências das xilose redutases obtidas. As barras coloridas a verde indicam aqueles organismos cujas XR possuem mutações, por exemplo, na <i>Candida tenuis</i> , foram encontradas duas sequências, a da sua xilose redutase e uma mutação dessa enzima.	25
Figura 12 - Gráfico dos K _M em mM da enzima xilose redutase WT (azul) e mutada (verde) nos diferentes organismos. A mutação escolhida para a <i>P.stipitis</i> (<i>Scheffersomyces stipitis</i>) foi K270S/N272P/S271G/R276F e para o organismo <i>Rasamsonia emersonii</i> foi a K271R/N273D.	26
Figura 13 - Alinhamento de todas as sequências das xilose redutases com os seus respetivos tamanhos. Realçado a vermelho aparecem aminoácidos comuns nos centros ativos de todas as enzimas e intervenientes na reação de redução da xilose (A), usando o clustal omega [123]. Árvore filogenética das sequências xilose redutases dos vários organismos e respetivos scores das distâncias (B).....	29



Figura 14 - Estrutura cristalizada da xilose redutase dimérica com o cofator NAD ⁺ do organismo <i>Candida tenuis</i> (1JEZ) [92].	30
Figura 15- Sobreposição de todas as estruturas da xilose redutase com o NAD representado a azul claro.	30
Figura 16 - Gráfico circular do número de enzimas XDH (incluindo a WT) e respectivos organismos a que pertencem.	31
Figura 17 - Gráfico que compara os diferentes K _M da xilose da <i>Scheffersomyces stipitis</i> e dos diferentes mutantes quando esta se encontra ligada ao NAD ⁺ (azul) ou NADP ⁺ (vermelho). Para alguns mutantes foi apenas conseguido encontrar dados para o K _M de um dos cofatores.	32
Figura 18 - Árvore filogenética das xilitol desidrogenases dos diferentes organismos e respectivos scores, usando a ferramenta clustal omega [123].	33
Figura 19 - Estrutura cristalizada da xilitol desidrogenase com o cofator NAD ⁺ do organismo <i>Gluconobacter oxydans</i> (1ZEM) usando o NGL viewer [124].	33
Figura 20 - Sobreposição de todas as estruturas secundárias da xilitol desidrogenases com o NAD representado a azul claro. Em ambas as figuras é possível visualizar o <i>fold</i> das proteínas.	34
Figura 21 - Estruturas em superfície da xilose redutase do <i>Aspergillus niger</i> (A, B) e do mutante K271R/N273D da <i>Rasamsonia emersonii</i> (C, D), ambas com os respectivos volumes dos centros ativos representados a laranja. Nas imagens A e C o NADH está embutido na estrutura a branco e está delimitado a vermelho o local onde se encontra o C4 do NADH, ao qual se vai ligar a xilose. Nas imagens B e D encontra-se delimitado a roxo o local onde o volume foi mal medido e transpõe o centro ativo.	36
Figura 22 - Gráfico da influência do volume das XR no seu K _M . Cada ponto azul representa o volume e respetivo K _M para cada uma das xilose redutases dos diferentes organismos e também os seus mutantes. Os volumes que foram mal delimitados foram descartados deste gráfico, obtendo a equação $y = -1,0625x + 261,01$ com um R ² de 0,3709.	37
Figura 23 - Gráfico da influência do volume das XR no seu K _M para os mutantes da levedura <i>Scheffersomyces stipitis</i> . Cada ponto azul representa o volume e respetivo K _M para cada uma das 3 xilose redutases mutadas deste organismo. O ponto verde corresponde à enzima não mutada. Este gráfico apresenta a equação $y = -0,8802x + 248,66$ com um R ² de 0,1063.	39
Figura 24 – Regressão linear entre os scores do <i>docking</i> e o volume. A equação da reta apresenta um declive negativo, sendo: $-0,0036x + 5,4203$, com um R ² de 2,872.	41



Figura 25 - Gráfico da relação entre o número de carbonos dos substratos da xilose redutase da <i>S.cerevisiae</i> e os seus respetivos K_M nesta enzima, com uma regressão linear $y=5,5739x - 8,9611$ e um R^2 de 0,2934. A vermelho encontram-se as pentoses e a verde os substratos com 3 carbonos. O ponto azul é referente à glucose (6C).....	43
Figura 26 - Docking molecular da L-arabinose (A), D-glucose (B), D-gliceraldeído (C), L-gliceraldeído (D) e da D-xilose (E) no centro ativo da xilose redutase do organismo <i>S.cerevisiae</i> . O tracejado amarelo representa os contactos polares com do substrato com as cadeias laterais dos aminoácidos da proteína.....	44
Figura 27 Centro ativo da xilitol desidrogenase do organismo <i>Rhizomucor pusillus</i> com o xilitol depois de ter sido sujeito ao processo de <i>docking</i> . É possível visualizar que os resíduos 300R, 156E e 52Y estabelecem ligações por pontes de hidrogénio (representadas a tracejado amarelo) com o xilitol.....	45



ÍNDICE DE TABELAS

Tabela 1 – Alguns programas de <i>docking</i> e respetivos algoritmos de procura.....	15
Tabela 2 - Tabela dos organismos a que pertencem as várias xilose redutases e valor dos respetivos volumes medidos pelo trj_cavity.	34
Tabela 3 - Os scores do docking da D-xilose nas diferentes xilose redutases e respetivos aminoácidos que estabelecem interações polares.....	40
Tabela 4 – Scores do <i>docking</i> de 5 diferentes açúcares com diferentes tamanhos na XR da <i>Saccharomyces cerevisiae</i>	42
Tabela 5 – Tabela com todas as xilitol desidrogenases e os seus respetivos scores de docking do xilitol juntamente com as interações polares que estabelecem no centro ativo. A abreviatura “pt” é referente a ligação por ponte de hidrogénio.	46



LISTA DE ABREVIATURAS, SIGLAS E ACRÓNIMOS

AKR – Aldo-Ceto redutase
Asn – Asparagina
Asp – Ácido aspártico
CO₂ – Dióxido de carbono
Dim – Dimensão
EA – Algoritmos Evolucionários
GA – Algoritmos genéticos
Gly – Glicina
His – Histidina
HXT – Transportadores de hexoses, do
IC – Construção incremental
LGA – Algoritmo genético Lamarckiano
Lys – Lisina
MC – Monte Carlo
MD – Dinâmicas moleculares
MDR – Desidrogenases/redutases de cadeia média
Min – Mínimo
NAD – Nicotinamida adenina dinucleótido
NADP – Nicotinamida adenina dinucleótido fosfato
NCBI - National Center for Biotechnology Information
NMR – Ressonância magnética nuclear
PDB – Protein Data Bank
PDH – Poliol desidrogenase
Pro – Prolina
SDH – Sorbitol desidrogenase
Tyr – Tirosina
WT – Wild-Type
XDH – Xilitol desidrogenase
XI – Xilose isomerase
XK – Xilulose quinase
XR – Xilose redutase



1. INTRODUÇÃO/ESTADO DA ARTE

A crescente taxa de exploração das reservas de petróleo e extração deste recurso, foram responsáveis por garantir todos os benefícios que podiam ser derivados do seu consumo ao longo vários anos, destacando-se o seu uso como combustível para os transportes. Até ao início do século 21, o facto deste recurso não renovável existir em abundância e ser barato, permitiu que muitas economias mundiais prosperassem quer pelo seu consumo, quer pela sua exportação[1,2]. Porém, as suas reservas começaram a esgotar e vão deixar de corresponder às necessidades da população mundial, apesar do seu consumo continuar a aumentar. Por este motivo, começaram a ser explorados outros combustíveis fósseis, que porém também já atingiram o limite da sustentabilidade. Contudo, os combustíveis continuam a ser a fonte de energia mais usada em todo o mundo. Por serem uma fonte de energia não renovável, o preço destes combustíveis continua a aumentar, bem como o impacto ambiental causado por estes. Os níveis de emissão de CO₂ e a sua consequente concentração na atmosfera terrestre têm vindo a crescer drasticamente desde o início da revolução industrial. A emissão de gases com efeito de estufa é o principal contribuinte para alterações climáticas no planeta, sendo o CO₂ um dos gases responsáveis por esse efeito, tendo como consequência principal o aquecimento global [1-4].

Por todos os motivos já referidos, a preocupação em encontrar soluções para diminuir o consumo dos combustíveis fósseis é cada vez maior e têm surgido novas soluções para contrariar esta tendência. Uma dessas alternativas passa pelo aproveitamento da biomassa lignocelulósica, do milho e da cana de açúcar para a produção de biocombustíveis [3]. Existe uma certa variedade de biocombustíveis; porém, é o etanol o que prevalece como o mais usado. O etanol é um álcool composto por 2 átomos de carbono, 1 de oxigénio e 6 de hidrogénio, cuja fórmula química é CH₃CH₂OH (figura 1). Existem vários fatores que contribuem para o etanol ser o biocombustível mais utilizado, tais como ter uma combustão bastante eficiente e um elevado valor octano [3].

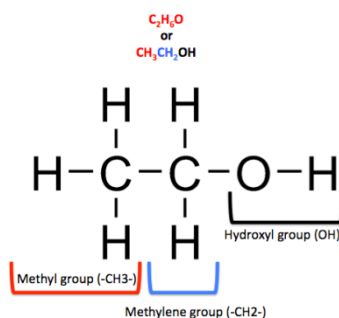


Figura 1 - Molécula de etanol e os seus respetivos grupos funcionais: colorido a vermelho o grupo metil (CH_3), um grupo metileno (CH_2) a azul e um grupo hidroxilo (OH) a preto [5].

É a partir desses recursos ricos em hidratos de carbono que vão ser obtidos os açúcares que vão funcionar como substrato para a produção de etanol, os quais provêm dos principais constituintes da biomassa lignocelulósica, sendo esses constituintes a celulose, hemicelulose. Os dois primeiros componentes, ao serem hidrolisados, libertam os seus açúcares constituintes, tanto hexoses (açúcares de 6 carbonos), sendo a glucose o açúcar mais abundante, como pentoses (açúcares de 5 carbonos), provenientes da hemicelulose, das quais se destacam a D-xylose e a L-arabinose, representando mais de 20% dessa biomassa (figura 2). A conversão destes açúcares em etanol é conseguida através dum processo designado por fermentação [2,3,4,6]. Esta é a segunda fase do processo de obtenção do bioetanol a partir da cana do açúcar, biomassa lignocelulósica e milho, sendo também esta a fase em que nos vamos focar neste trabalho.

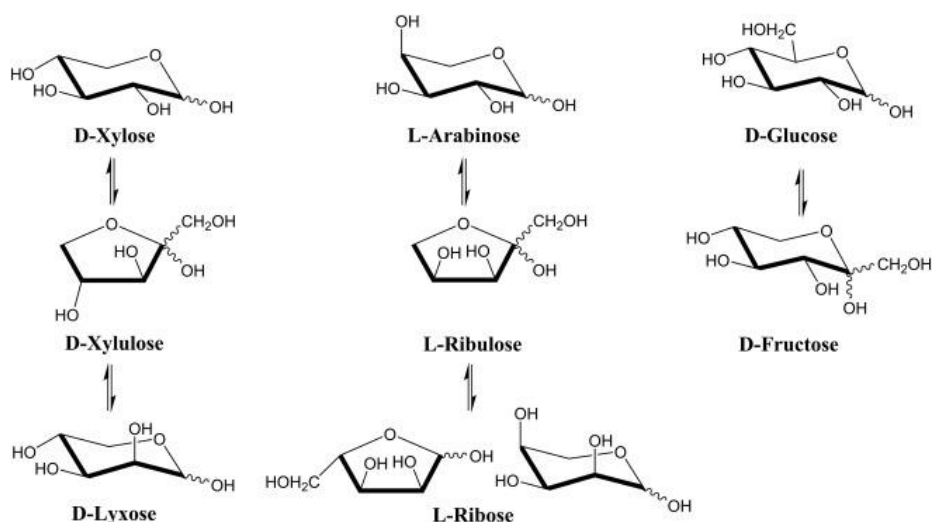


Figura 2 - Estrutura de alguns dos açúcares obtidos a partir da hidrólise da biomassa lignocelulósica e algumas das suas possíveis reações [7].



O processo de obtenção de etanol através duma via fermentativa é conseguida através da utilização de microrganismos, sendo a *Saccharomyces cerevisiae* o mais utilizado para fins industriais. A *S. cerevisiae* é um eucariota unicelular, mais propriamente uma levedura pertencente ao reino dos fungos, e que, por esse motivo contém organelos membranares como núcleo, mitocôndria, entre outros [8]. A utilização desta levedura deve-se sobretudo a fatores como a elevada tolerância a baixos níveis de pH e elevados níveis de etanol, resistência a bacteriófagos e também por possuir a capacidade de fermentar açúcares em condições anaeróbicas bastante severas. Este microorganismo possui ainda um elevado rendimento e grandes taxas de produção de etanol a partir das hexoses como a glucose, um processo que está bem definido a nível industrial [2,3,9]. Porém, os açúcares obtidos através da hidrólise, para além de conterem hexoses, também possuem uma porção significativa de pentoses provenientes das hemiceluloses, destacando-se a D-xilose como pentose mais abundante. É deste fator que advém o primeiro problema da utilização da *S. cerevisiae*, uma vez que esta levedura não tem capacidade de utilizar pentoses para a produção de etanol, ficando assim cerca de 13-39% da matéria desaproveitada, problema que apenas foi abordado por volta do ano 1970 [3,4,9]. A *S. cerevisiae*, apesar de não ter a capacidade de crescer num meio com apenas xilose, consegue fermentar diretamente a D-xilulose, apresentando contudo uma taxa de conversão 10 vezes inferior à sua taxa de conversão de glucose [4]. Isto, aliado a outros problemas como alguns dos produtos resultantes (furaldeídos, acetato, compostos fenólicos...) da primeira fase de hidrólise dos constituintes da biomassa lignocelulósica inibirem a fermentação, constituem fortes entraves que têm de ser ultrapassados [3].

1.1 Modificação genética de *Saccharomyces cerevisiae*

Grande parte das leveduras possui a capacidade de fermentar xilose em etanol, tais como a *Pachysolen tannophilus* e a *Candida tropicalis*. Porém, a *Saccharomyces cerevisiae* não é uma delas. Esta possui os „open reading frames“ que codificam para as três primeiras enzimas envolvidas no metabolismo da xilose: xilose redutase (XR), xilitol desidrogenase (XDH) e xilulose quinase (XK) [10, 11]. Contudo, os níveis de expressão dos genes que codificam a XR e XDH são de tal maneira baixos ou inesistentes que impossibilitam esta levedura de consumir a xylose de modo a conseguir sobreviver através deste substrato, sendo então necessário introduzir neste organismo genes heterólogos provenientes doutros microrganismos que possuam a capacidade de fermentar xylose, obtendo assim uma *Saccharomyces cerevisiae* recombinante [11]. Melhoramentos moderados na expressão da

XK são benéficos para estas *Saccharomyces cerevisiae* recombinantes [12]. A levedura *Scheffersomyces stipitis* metaboliza a xilose através da via descrita. Mas este microrganismo é sensível ao etanol e requer uma oxigenação baixa e cuidadosamente controlada, incapacitando o seu uso a um nível industrial, podendo então a via de fermentação da xilose deste organismo ser introduzida na *Saccharomyces cerevisiae*, clonando os genes codificadores da XR e XDH, XYL1 e XYL2 respectivamente, e introduzindo-os neste organismo [13,14].

Depois de introduzida a via XR/XDH da *Scheffersomyces stipitis*, a conversão de xilose em xilulose é conseguida por duas reações: primeiro, a xilose é reduzida a xilitol pela xilose redutase (XR) dependente do NAD(P)H e codificada pelo gene XYL1. A segunda reação é a oxidação de xilitol em D-xilulose pela xilitol desidrogenase (XDH) dependente do NAD⁺ e codificada pelo gene XYL2 [15,16]. Nas bactérias e alguns fungos, esta reação tem apenas um passo, através da xilose isomerase (XI), que consiste na outra via possível que pode ser introduzida em *S. cerevisiae* [17]. Após estas reações, a D-xilulose é convertida em D-xilulose-5P pela xilulose quinase (XK), com a D-xilulose-5P indo para a via das pentoses fosfato onde será convertida em etanol (figura 3) [15].

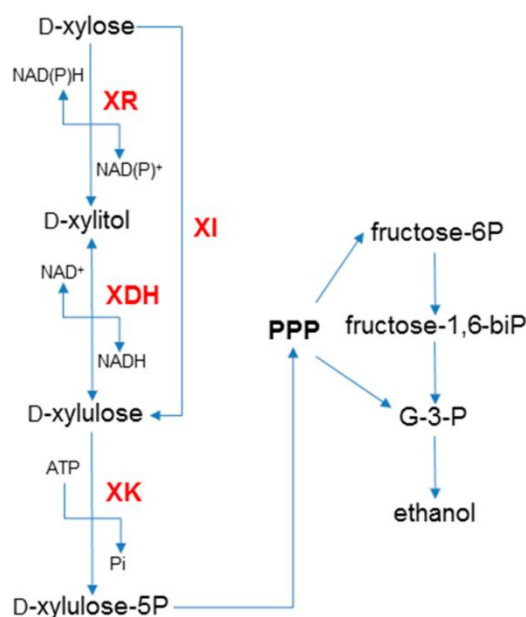


Figura 3 - A via de fermentação de xilose na *Saccharomyces cerevisiae* recombinante [3].

Assim, o objetivo deste trabalho é a análise da via metabólica da fermentação de xilose em *Saccharomyces cerevisiae* visando, através de abordagens *in silico*, descobrir na estrutura das proteínas e seus substratos intervenientes nesta via, fatores que possam influenciar o aproveitamento ineficiente da xilose para a obtenção de etanol, focando sobretudo nos açúcares envolvidos nas reações destas enzimas. Com este fim, uma das características



estruturais analisadas foi a medição dos volumes dos centros ativos para serem comparados com parâmetros de medida de afinidades como o K_M . Diferentes volumes originam diferentes espaços e modos possíveis de ligação, pelo que prever o local, o modo de ligação e a conformação da xilose, xilitol e outros substratos foi um dos objetivos, que entre os quais se incluem:

- Recolher informações de várias XR e XDH de diversos organismos e de dados experimentais relativos a essas enzimas.
- Construir um modelo das enzimas recolhidas que ainda não tivessem uma estrutura na PDB.
- Investigar qual o volume dos centros ativos das enzimas recolhidas e a sua correlação com o K_M ou os resultados *docking*.
- Analisar algumas interações que possam ocorrer no centro ativo depois dos substratos serem sujeitos ao *docking* molecular.
- Perceber eventuais falhas e vantagens nos métodos e programas usados para medir os volumes e para o *docking*.

1.2 Problemas com a *Saccharomyces cerevisiae* modificada com a via XI e XR/XDH e soluções já exploradas

1.2.1 Modificação dos genes que codificam os transportadores de hexoses para melhorar a entrada e consumo da D-xilose

A levedura *Saccharomyces cerevisiae* consegue crescer num meio de elevadas concentrações de glucose pois existem genes que codificam transportadores de hexoses (HXT) fazem com que o consumo de D-glucose seja facilitado [18]. Para que o processo de fermentação seja industrialmente rentável é necessário converter a D-xilose e D-glucose de forma rápida e eficiente em etanol por parte das estirpes modificadas de *Saccharomyces cerevisiae*, pois existem alguns efeitos inibitórios causados por metabolitos no metabolismo da D-xilose [19,20]. Com este fim, foi usada uma estirpe DS68616 [21] que metaboliza a D-xilose de forma eficiente, mas em contrapartida, na presença de elevados níveis de glucose, o transporte de xilose é inibido [21]. O açúcar preferencial dos transportadores HXT é a glucose, sobrepondo-se assim à xilose, fazendo com que esta não seja consumida. Por isso, foi desenvolvida outra estirpe, a DS69473 [21], cuja via de degradação da xilose foi a da xilose isomerase dos fungos e sofreu uma quadrupla deleção das hexokinases (GLK1, HXK1, HXK2 e GAL1), que visava ultrapassar este desafio; porém, era incapaz de fermentar a glucose apesar da xilose ser eficientemente consumida [21]. Esta estirpe foi então submetida durante 40 dias a um crescente aumento da concentração de glucose, à qual a estirpe se



adaptou, aumentando assim a percentagem de consumo de glucose comparativamente à xilose, da qual surgiu uma estirpe evoluída, a DS69473Evo. Esta levedura é capaz de fermentar elevados níveis de xilose e glucose em simultâneo, sendo até capaz de melhorar o consumo de xilose à medida que as concentrações de glucose vão subindo. Por exemplo, com uma concentração de xilose de 2% e de glucose 12%, a DS69473 era incapaz de crescer enquanto a DS69473Evo mostrou um crescimento bastante elevado. Contudo, parte do consumo da D-xilose continua a ser inibido e para níveis muito altos de glucose o seu crescimento continua a ser inibido [20,21].

1.2.2 A acumulação de xilitol e trocas de afinidades dos cofatores

A solução mais utilizada para tornar a *S. cerevisiae* num microrganismo capaz de fermentar xilose é a incorporação dos genes heterólogos da *Scheffersomyces stipitis* que codificam as enzimas XR e XDH [16,22]. Porém, há uma acumulação e excreção de xilitol resultante da primeira reação, o que constitui um obstáculo para tornar a utilização da *S. cerevisiae* recombinante num processo industrialmente rentável [22,23]. Uma das razões para isto se suceder é o desequilíbrio redox entre as enzimas XR e XDH. A xilose redutase utiliza preferencialmente o NADPH como cofator, o que não se sucede com a xilitol desidrogenase, que usa apenas o NAD⁺, levando assim a um desequilíbrio entre cofatores e consequentes perdas de capacidades de assimilação de carbono e de produção de etanol [24,25]. Para resolver este problema têm sido implementadas diversas estratégias de engenharia metabólica como conectando a redução do furaldeído com o metabolismo da xilose [26], canalizando os fluxos de carbono através de uma via de fosfocetolase numa estirpe recombinante que consome xilose [27], alterando a via de assimilação da amónia [28], adicionando um aceitador externo de eletrões aos meios de fermentação [29] e alterando o cofator preferencial[24].

Já foi demonstrado em estirpes de *S. cerevisiae* metabolizadoras de xilose que, por exemplo, com a deleção do gene *zwf1*, gene codificador da glucose-6-fosfato desidrogenase, esta consegue aumentar significativamente a produção de etanol de 0.31 g g⁻¹ para 0.41 g g⁻¹ acompanhado de uma diminuição da produção de xilitol [30]. Isto porque a reação catalisada por esta enzima era responsável pela redução do NADP⁺ a NADPH na via das pentoses fosfato, ficando assim este cofator indisponível para ser utilizado pela enzima xilose redutase, incapacitando assim esta enzima de reduzir a xilose em xilitol pelo NADPH proveniente da via não oxidativa das pentoses fosfato, favorecendo o uso de NADH nesta reação [28,30].



Em relação a uma alteração dos cofatores, Khattab et al.[31, 32, 33] fizeram uma “site-directed mutagenesis” de modo obter uma enzima XR modificada que fosse apenas dependente do NADPH, usando a XR do microrganismo *Scheffersomyces stipitis*, tendo sido obtidos 3 mutantes diferentes. Os mutantes E223A/S271A (AA), E223D/S271A (DA), e E223G/S271A (GA) conseguem manter uma atividade catalítica da enzima semelhante à não mutada, usando apenas o NADPH como cofator. No caso do duplo mutante E223A/S271A, a atividade enzimática foi de 106% comparada com a original. Novamente, comparando com o original, o duplo mutante E223D/S271A mostrou um aumento da atividade de 1.27 vezes [31,32,33]. Outros autores seguiram uma diferente abordagem. Em vez de tornar a XR mais específica para o NADPH, Watanabe et al. [24] propuseram tornar a enzima XDH específica para o cofator NADP⁺ em vez do NAD⁺, ficando com uma Wild-Type XR e uma XDH mutada provenientes na mesma do organismo *Scheffersomyces stipitis* [24]. Estes autores também já tinham construído uma XR específica para o NADPH [25]. Os resultados relativamente à XDH foram notórios, tendo-se verificado uma redução do xylitol acumulado de 86% juntamente com um aumento de 41% na produção de etanol por parte desta *S. cerevisiae* recombinante mutada. [24, 33]. A mesma estratégia também foi testada com a sobre expressão da xilulose quinase (XK) endógena, o que resultou numa diminuição de 28% da acumulação do xilitol e um aumento da produção de etanol em 15% [29]. Porém, ao aplicar esta segunda abordagem, continua a haver um desequilíbrio redox devido ao uso dos dois cofatores na atividade catalítica da XR [32,33].

1.3 Biologia estrutural

A análise das estruturas tridimensionais das proteínas permite a obtenção de uma maior informação acerca da sua função comparativamente a uma análise de apenas a sua estrutura primária. Como a sua sequência pode ser facilmente determinada pela sequência do gene que a codifica e também devido às novas tecnologias de sequenciação (“high throughput next-generation sequencing”), as sequências que entram nas bases de dados são cada vez mais, um aumento muito significativo como é possível verificar na PDB e Swiss-Prot [34,35]. Esta última base de dados em 1986 tinha apenas 3939 sequências, numero que aumentou de forma exponencial até 2019 onde existem cerca de 550.000 sequências guardadas [36,37,38]. Em 2019 a base de dados PDB tinha cerca de 147.000 estruturas já cristalizadas guardadas, um



número bastante menor do que as sequências já existentes na Swiss-Prot [39,40]. Face a esta discrepância de números, era necessário arranjar métodos para entender os mecanismos de funcionamento das proteínas que ainda não possuíam uma estrutura tridimensional definida. Apenas deste modo seria possível identificar as reações em que estavam envolvidas, a sua localização sub-celular, características evolucionárias, entre outros atributos, sendo tal possível apenas com um modelo 3D da proteína [34, 35, 41]. Estas estruturas são determinadas a nível experimental sobretudo por duas técnicas, cristalografia raio-X e mais recentemente espectroscopia de ressonância magnética nuclear (RMN), técnicas que apesar de precisas, são bastante demoradas, sendo que daí advém o problema de haver demasiadas sequências de proteínas para poucas estruturas tridimensionais [42]. Para diminuir a discrepância entre estes números (sequências e estruturas), a modelação baseada em homologia é atualmente a abordagem mais confiável, rápida e económica para deduzir as propriedades estruturais das sequências, porém salientando que estes modelos de proteínas computacionais só podem ser validados experimentalmente [42,43]. A criação de modelos por homologia pode ser realizada fornecendo um alvo (target), uma sequência dum proteína e alinhar esta sequência a uma ou várias sequências cujas estruturas já são conhecidas (templates) [42-45]. Um alinhamento é considerado correto e de qualidade quando a identidade entre o target e o template é superior a 40%. Valores inferiores a este diminuem o grau de fiabilidade do alinhamento pois este torna-se mais complicado, o que diminui também a qualidade do modelo [43,45].

Existem vários programas usados para realizar a modelação de proteínas por homologia, tais como o HHsuite [46], o MolIDE [47], o RosettaCM [48], o SCWRL [49] e o Modeller [50], sendo este último o usado neste trabalho. O Modeller usa, no seu caso mais simples, três inputs: o alinhamento entre uma estrutura dum proteína e a sequência doutra, e as coordenadas atómicas do template para calcular o modelo [50].

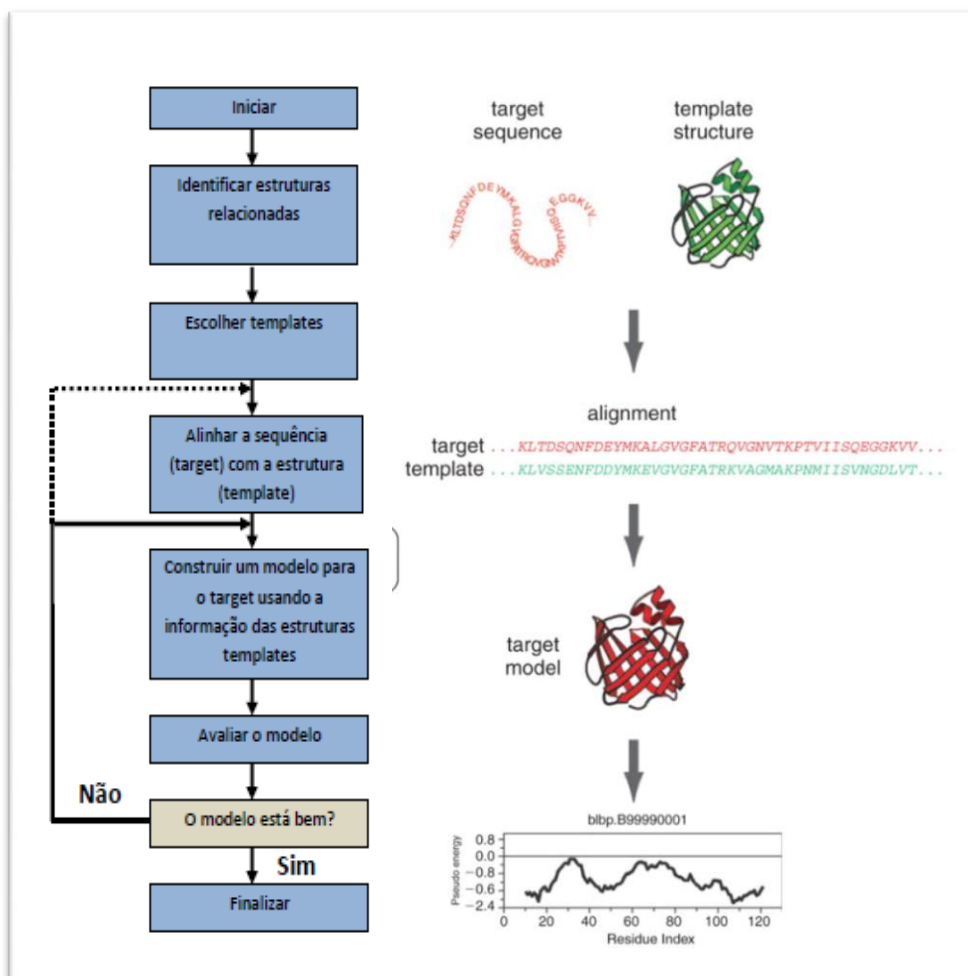


Figura 4 - Esquema do processo de modelação de proteínas por homologia [50].

1.3.1 Centro ativo de Enzimas

Há mais de um século atrás Emil Fischer observou que a ligação de um substrato a uma proteína se assemelha à inserção de uma chave numa fechadura [51]. Com base nesta descoberta, os métodos para identificar os centros ativos de enzimas giram à volta deste pressuposto, pois há uma complementaridade entre as formas do local ativo da enzima e do substrato. Assim, uma das características destas macromoléculas é que possuem várias cavidades e túneis associados à sua estrutura terciária e quaternária [52,53]. Estas cavidades são superfícies côncavas, vazias, e geralmente encontram-se no cerne das proteínas, podendo também existir numa parte mais superficial destas [54]. Geralmente o centro ativo é a maior cavidade duma proteína, sendo verificado em 83% dos casos em estudos realizados por Laskowski et al [55]. Estes “pockets” criam ambientes específicos dentro das enzimas com propriedades físico-químicas próprias, como o pH e a força iónica, para facilitar a catálise enzimática, através de zonas de alta afinidade e especificidade para os ligandos. A identidade



e disposição dos aminoácidos no centro ativo definem a topologia das enzimas em termos de estereoquímica, hidrofobicidade e carácter eletrostático, características que podem ser analisadas computacionalmente como será descrito na próxima secção e que determinam quais os substratos que se vão ligar à enzima [53].

Existem 20 aminoácidos naturais sendo que apenas 11 deles são geralmente encontrados diretamente ligados à atividade catalítica duma enzima, tendo 7 diferentes grupos químicos. A histidina possui um grupo imidazola, a lisina um grupo amina, a glutamina e a asparagina um grupo amida, a cisteína um grupo tiol, a serina, treonina e tirosina um grupo hidroxilo, a arginina um grupo guanidínio e o glutamato e aspartato possuem um grupo carboxilato [56]. Cada cadeia lateral tem o seu pK_a que pode ser alterado com certas interações de alguns grupos químicos, geralmente da seguinte forma: quando um grupo entra em contacto com uma carga negativa, o pK_a aumenta, assim como o contrário também se verifica: quando em contacto com uma carga positiva o pK_a diminui. Para além destes aminoácidos as enzimas também usam iões metálicos, cofactores e moléculas de água no seu processo de catálise [57].

1.3.2 Métodos computacionais para identificação de centros ativos e análise das suas propriedades

Existem diversos métodos computacionais para identificar os locais ativos onde ocorre a ligação entre o ligando e a proteína. Estes são importantes pois permitem prever qual o local ativo de uma proteína não caracterizada. Esses métodos podem ser classificados em puramente geométricos, energéticos, sendo que estes dois métodos consideram a estrutura da proteína como estática, e métodos baseados noutros critérios como: comparação entre estruturas e sequência e também métodos que têm em consideração a dinâmica da estrutura das proteínas [53].

Os métodos geométricos para deteção e identificação de locais ativos têm por base que o maior pocket da proteína é aquele onde está localizado o local ativo. São usados vários programas com diferentes algoritmos que usam métodos geométricos, nos quais estão incluídos o POCKET[58], o CAST[59], o PASS[60] e o LigandFit[61] sendo que nenhum destes métodos requer qualquer informação e conhecimento sobre os ligandos [62].

Nos métodos energéticos é usado o princípio de que o centro ativo é o local onde há o mínimo de energia [63, 64]. Para uma abordagem usando um critério energético, o programa Q-SiteFinder [65] calcula as forças de van der Waals através de sondas de grupos metil dentro



da proteína. Essas sondas envolvem a proteína e vão procurar as interações de baixa energia, onde se tiverem detetado uma interação com baixa energia, guardam-na. Depois estas são agrupadas por clusters, que por sua vez são classificados de acordo com o número de sondas no cluster, sendo que aquele com maior número de sondas ou que for energeticamente mais favorável é considerado o primeiro classificado e também aquele que é mais provável ser o centro ativo da proteína [65]. O AutoLigand [66] e o Grid [64] são outros exemplos de outros algoritmos que se baseiam numa abordagem energética.

Outro tipo de métodos computacionais são os de comparação dos locais ativos. Estes vêm complementar os métodos convencionais que usam a informação da sequência e do fold da anotação das estruturas das proteínas [63]. Um dos exemplos desses métodos é o IsoCleft. O IsoCleft não requer nenhuma informação sobre as diferentes ligações que os ligandos podem estabelecer na proteína nem um alinhamento da sequência, mesmo sendo um método que procura semelhanças 3D de proteínas emparelhadas [67]. Outro método de descrição e comparação de locais ativos é o Cavbase, comparando locais ativos de diferentes proteínas através das suas propriedades físico-químicas e geométricas [68]. ConSurf [69], Rate4Site [70] e FINDSITE [71] são exemplos de algoritmos que usam um conhecimento prévio das funções das proteínas de modo a identificarem o local ativo.

Todos os métodos descritos acima correspondem a abordagens que não têm em consideração uma eventual mudança conformacional da proteína ao longo do tempo, o que acontece em muitos fenómenos biológicos. Por vezes esses métodos não conseguem detetar as cavidades que só ficam acessíveis em conformações específicas. Para colmatar este problema, têm sido desenvolvidos vários métodos que usam simulações de dinâmicas moleculares (MD) para análise das cavidades através dos vários estados conformacionais das proteínas. Exemplos de programas que usam MD são: Caver3.0 [72], Epock [73], dxTuber [74], MDpocket [75] e o trj_cavity [52], sendo este último o usado neste trabalho.

O programa trj_cavity identifica e analisa de uma forma rápida e eficiente todas as cavidades (pockets, túneis). Foi desenvolvido por Paramo et al. e tem por base uma nova abordagem “grid-based” que geralmente tem um melhor desempenho do que aqueles que usam uma procura exaustiva por todos os átomos emparelhados, o que gasta muito mais recursos do computador e tempo devido ao número elevado de operações deste processo. O algoritmo de procura de cavidades do trj_cavity usa voxels, em que se cada voxel estiver vazio vai procurar o voxel vizinho numa das 6 direções possíveis dos eixos x, y e z. Sempre que encontra um voxel vazio continua a procura até encontrar a superfície da proteína, onde depois recua para a lista que guardou dos voxels vizinhos vazios e continua até encontrar a

superfície da proteína em cada caminho possível. Quando não houver mais voxels vazios a visitar, a cavidade foi encontrada e pode procurar a próxima (figura 5) [52,62].

Os resultados deste programa estão fortemente dependentes dos parâmetros que o utilizador lhe fornece, como o tamanho do voxel (spacing), o volume mínimo da cavidade a detetar (min), o número de direções de procura do voxel (dim), onde por exemplo nos túneis é mais provável de requerer um dim de 4 mas se for uma cavidade mais interna já requer provavelmente as 6 direções [52, 62].

Depois de detetados, os centros ativos das proteínas têm propriedades físico-químicas que podem ser analisadas computacionalmente. A composição de aminoacídica dos centros ativos pode ser detetada através das ferramentas ConSurf e InterPare [76]. Em relação à hidrofobicidade existem programas como o HINT [77], o SuperStar [78], ou o MLP [79] que calculam a hidrofobicidade quer na proteína, quer no seu centro ativo. As análises geométricas, como a medição do volume, tamanho e superfície do centro ativo permitem fazer uma primeira análise ao centro ativo. O SCREEN [80] e o CAST estão na lista de programas capazes de analisar algumas propriedades geométricas. Já o trj_cavity consegue por exemplo indicar qual o volume e superfície da cavidade que queremos procurar [52].

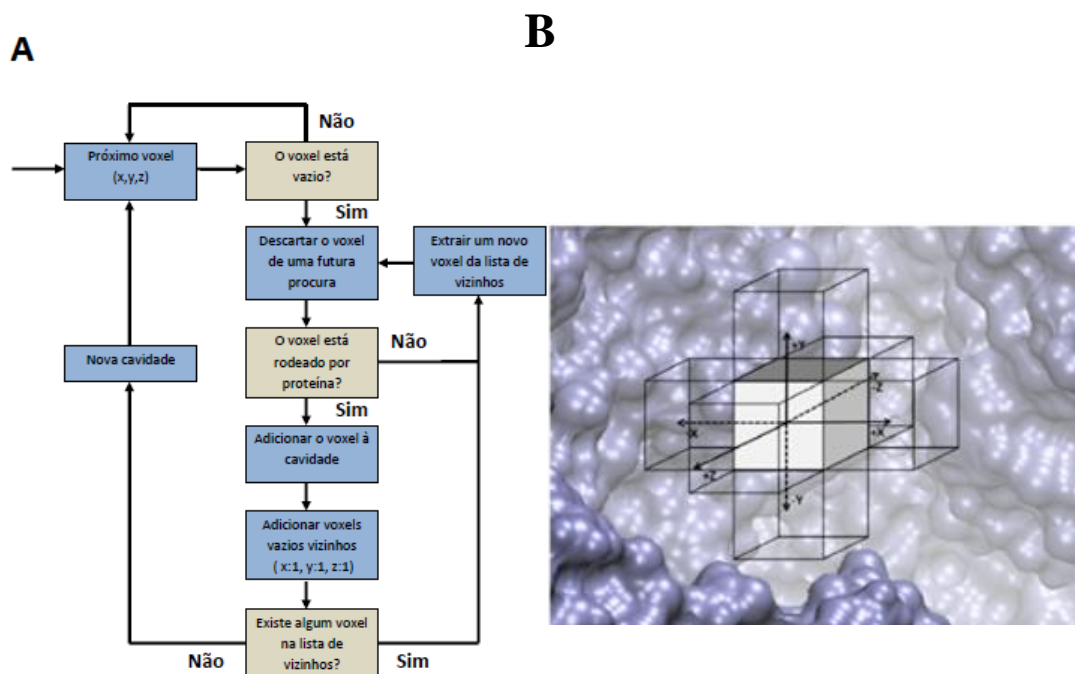


Figura 5 - Metodologia de procura do algoritmo usado pelo trj_cavity (A) e representação de um voxel e as suas possíveis direções de procura (B) [52].

1.3.3 Molecular Dynamics e MD simulations

Um dos ramos em crescimento da biologia de sistemas é estudar e prever as interações nas proteínas. Ter uma estrutura tridimensional de uma proteína, quer seja das armazenadas na PDB ou conseguida por homologia, é crucial mas não conseguem providenciar as dinâmicas moleculares a que estão sujeitas durante alguns processos biológicos nem as interações entre as várias moléculas. Tanto as proteínas como os seus ligandos são estruturas flexíveis, sofrendo alterações conformacionais e rearranjos na sua estrutura durante o processo de interação com este, como por exemplo durante uma reação química ou no caso da proteína, dependendo do substrato ligado a ela (figura 6), sendo então crucial desenvolver métodos de dinâmicas moleculares [81].

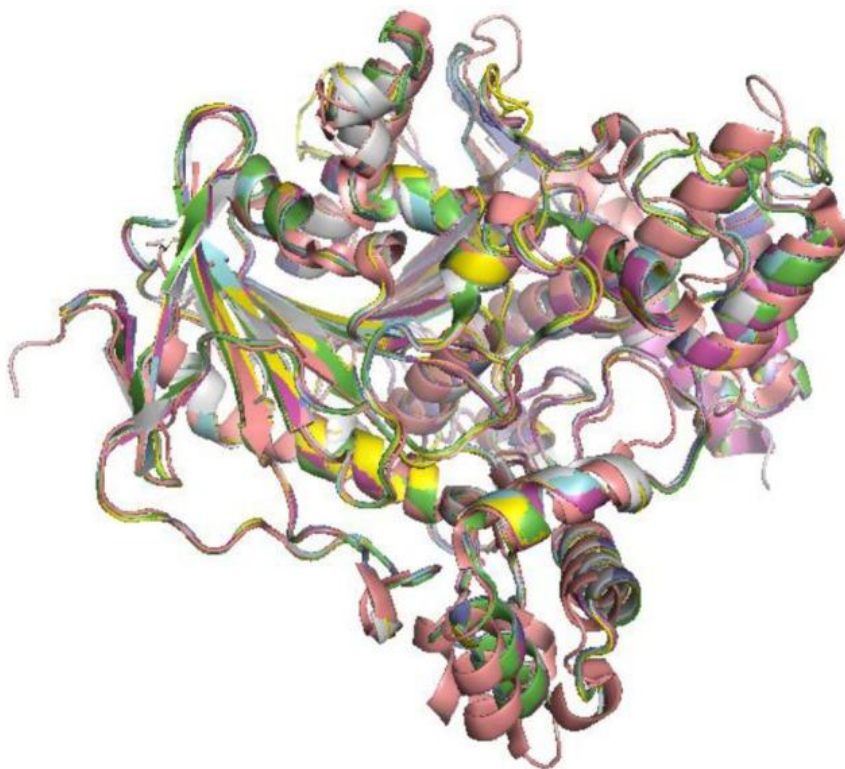


Figura 6 - Exemplo das diferentes conformações da mesma família de proteínas sobreposta mas com diferentes substratos ligados [81].

1.3.4 Molecular Docking

Os métodos de *docking* têm sido dos mais bem sucedidos no processo de planeamento e design de novos compostos, tendo como objetivo prever e racionalizar diferentes interações químicas e biológicas entre as várias moléculas sujeitas a estes métodos [82,83]. Geralmente este método é usado para prever e classificar as conformações e afinidades das ligações entre uma macromolécula e uma molécula pequena, como um ligando, em estruturas que foram



geradas por simulações MD ou por modelação por homologia [83,84]. Contrariamente aos métodos de *molecular dynamics*, os métodos de *docking* para determinar as conformações das moléculas não lidam diretamente com as energias e potenciais químicos. Estes assumem a proteína (recetor) como um corpo rígido, não alterando a sua conformação entre o estado de quando se encontra ligada ao ligando ou não, com ângulos e tamanhos das ligações covalentes constantes mas com algumas ligações previamente definidas a terem uma rotatividade total [82,85].

Os métodos de *docking* começaram quando Levinthal et al., há mais de 40 anos atrás, usaram esta metodologia para prever possíveis conformações das fibras de hemoglobina [86]. Desde essa data que os métodos de *docking* têm ficado cada vez mais desenvolvidos, aumentando a precisão das previsões e diminuindo o tempo computacional gasto, sendo que já existem mais de 60 programas e 30 funções de avaliação (*scoring*) destinados a este fim [87]. Entre estes 60 programas são apenas 9 os mais utilizados, sendo eles o AutoDock [88], DOCK [89], FlexX [90], Glide [91], GOLD [92], ICM [93], QXP [94], Surflex [95] (tabela 1). Todos estes programas têm dois aspetos principais: o método de exploração da conformação espacial do ligando ou proteína (o algoritmo), dependendo de qual o escolhido como alvo, e uma função de avaliação que irá avaliar se a conformação espacial proposta é a mais correta. Este segundo aspeto é conseguido através da atribuição do melhor *score* a uma dada posição do ligando na estrutura cristalizada, servindo de base para o funcionamento do algoritmo. Numa visão mais química, a classificação dos ligandos pela função de avaliação é um resultado das interações por pontes de hidrogénio, forças de Van der Waals, forças repulsivas e parâmetros de solvatação. Aos ligandos mais ativos na proteína também lhes é atribuído um melhor *score* [87,96,97,98].

**Tabela 1** – Alguns programas de *docking* e respectivos algoritmos de procura.

Programa	Algoritmo de procura
AutoDock 4.0	Algoritmo genético (GA)
Dock	Construção incremental (IC)
FlexX	Construção incremental (IC)
Surflex	Construção incremental (IC)
GEMDock	Algoritmo evolucionário (EA)
GOLD	Algoritmo genético (GA)
ICM	Monte-Carlo (MC)
SLIDE	Construção incremental (IC)

Diferentes métodos de amostragem computacional têm sido usados por estes programas, como os algoritmos *Multiconformer docking*, métodos estocásticos e *consensus docking*, sendo o segundo o mais focado neste trabalho [99]. Assim, em relação aos métodos estocásticos, eles têm por base a exploração da conformação do ligando, sendo usados geralmente dois métodos estocásticos de procura, os Algoritmos Genéticos (GAs) e a procura Monte Carlo. Os programas de *docking* que usam GAs são por exemplo o Autodock, o GOLD e o FITTED [98]. Já programas como o DockVision ou o ProDOCK usam o Monte Carlo [87]. Os GAs seguem a teoria da evolução de Charles Darwin. Cada ligando corresponde a um cromossoma. Cada cromossoma tem genes, esses genes codificam por sua vez ângulos de torção, translação e rotação do ligando no espaço, sendo que a conformação que passa para a geração seguinte é aquela que apresenta um score menor (menor energia) (figura 7) [99].

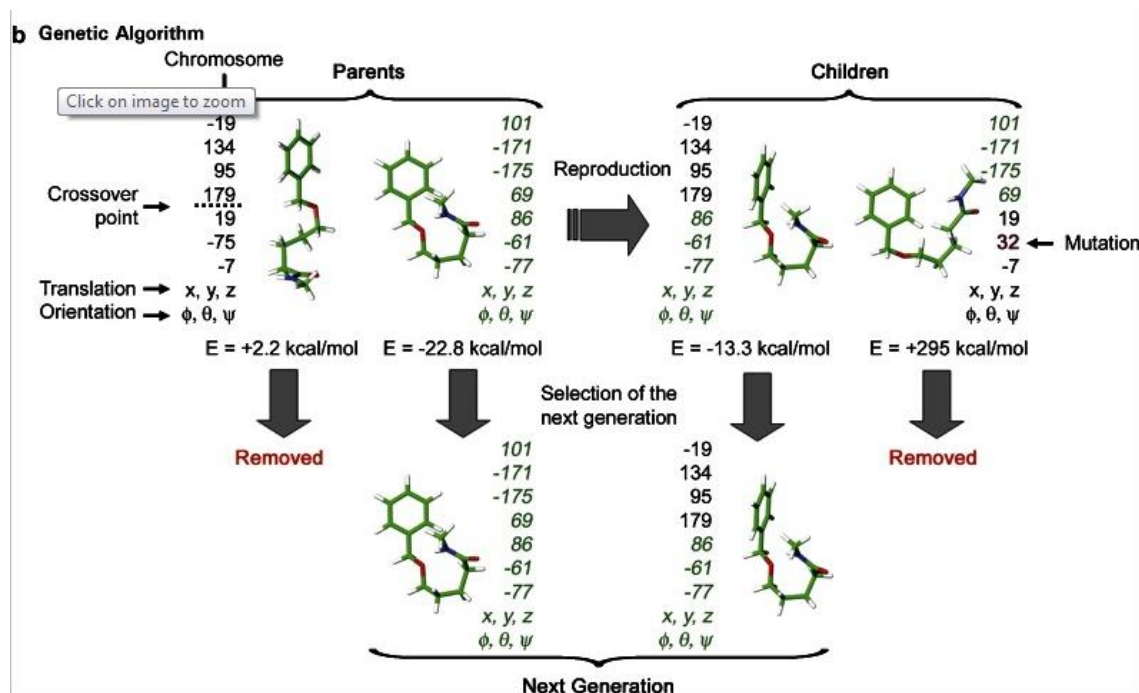


Figura 7 - Mecanismo de procura dos GAs [99].

No programa AutoDock a estratégia base de *docking* entre um ligando e proteína começa a partir do exterior da proteína, em que aleatoriamente se testam diferentes conformações, translações e orientações do ligando até encontrar o local ideal na proteína. Assim, o AutoDock utiliza o chamado docking flexível, sendo por isso bom para arranjar a melhor ligação entre uma pequena molécula e uma macromolécula, por exemplo um ligando e uma enzima. O método de pesquisa mais eficiente no AutoDock, aquele que geralmente dá resultados com menor valor de energia, é o Lamarckian Genetic Algorithm (LGA), uma das várias modificações dos GAs [83,96,100].

Um dos programas de *docking* mais usados em todo o mundo e que foi desenvolvido pelo mesmo grupo que criou o AutoDock é o AutoDock Vina. Lançado em 2009, este permite alguma flexibilidade das moléculas pois confere às moléculas alguns graus de liberdade, não sendo um corpo rígido, como por exemplo algumas cadeias laterais de aminoácidos previamente definidas [101].

Têm sido lançadas versões mais recentes e melhoradas do programa AutoDock Vina como o QuickVina [97], PSOVina [102] e o AutoDock VinaXB [103].

1.3.5 As duas enzimas principais da via de fermentação da xilose

Após terem sido descritas algumas abordagens metabólicas e os métodos *in silico* usados na biologia estrutural, resta saber quais as características estruturais e estudos que já

foram realizados sobre as duas enzimas que serão alvo de estudo neste trabalho: a xilose redutase (XR) e a xilitol desidrogenase (XDH).

A xilose redutase é uma das enzimas pertencentes à superfamília das aldo-keto redutases (AKR) [104]. Quase todas as AKR utilizam o NAD(H) e o NADP(H) como um dos cofatores nas suas reações, sendo que são oxidoreductases que desempenham funções nos mais diversos seres vivos, desde vertebrados, invertebrados e plantas, a protozoários, fungos, eubactérias, a arqueobactérias. Esta superfamília, com aproximadamente 105 membros, apresenta-se dividida em 14 famílias. Estas geralmente são proteínas (α/β)8-barril com um tamanho de cerca de 320 aminoácidos, sendo que todas elas apresentam na sua estrutura primária 11 aminoácidos em certas posições, sendo eles: Gly-22, Gly-45, Asp-50, Lys-84, Asp-112, Pro-119, Gly-164, Asn-167, Pro-186, Gln-190 e Ser-271 (figura 8). Estas contêm um centro ativo com um tetraedro catalítico composto por uma tirosina, uma lisina, um aspartato e uma histidina [104,105,106]. Entre dois destes resíduos, tirosina e histidina, juntamente com o C-4 do NADP⁺ quando este cofator se encontra ligado, geralmente é encontrada uma molécula de água ligada, sendo que a sua função está provavelmente relacionada com um arranjo que se pensa mimetizar a localização do substrato. Outra característica do arranjo espacial deste centro ativo é a criação de um espaço oxianião entre a Tyr, His e o anel do NAD(P). Como é no carbono 4 do anel de nicotinamida do NAD(P) que ocorre a protonação desta molécula, os aminoácidos Tyr ou His serão os dadores do próton devido à sua proximidade com o C-4 e os respetivos grupos carbonilo destes dois resíduos. Relativamente aos outros dois aminoácidos presentes no local ativo pensa-se que servem para estabelecer ligações por pontes de hidrogénio de modo a facilitar a cedência do H⁺ por parte da tirosina, pois há uma diminuição do seu pK_a [106].

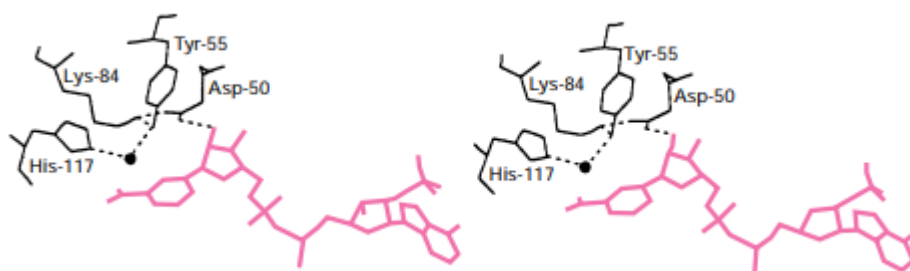


Figura 8 - Disposição espacial dos resíduos Asp-50, Tyr-55, Lys-84 e His-117 presentes no centro catalítico das AKR (subfamília da Xilose Reductase) e as suas interações com a molécula de água (preto) e o anel de nicotinamida do NAD (rosa) [106].



De um modo geral, estas enzimas privilegiam substratos hidrofóbicos em relação aos hidrofílicos. Estudos demonstraram que esta enzima segue um mecanismo bi-bi, sendo que o substrato, neste caso a D-xilose, é a última a ligar-se, sendo o cofator o primeiro e que por sua vez também é o último a sair do centro ativo. A XR é uma enzima que, tal como a maioria das AKR, pode catalisar a transformação de uma aldopentose num álcool, que é a reação preferencial, embora esta seja reversível [104-107].

Relativamente à XDH, esta é uma das enzimas pertencentes à superfamília de enzimas desidrogena/redutase de cadeia média (*medium-chain dehydrogenase/reductase* - MDR). Uma característica comum a todas as enzimas MDR é que elas utilizam o NAD(H) ou NADP(H) como cofator. De realçar que também, a grande maioria delas possuem um íon zinco com uma função catalítica no centro ativo. Dentro desta superfamília (constituída por cerca de 1000 enzimas) existem 8 subfamílias, cuja divisão foi feita de acordo com a semelhança estrutural dos substratos e do alinhamento da sequência dos aminoácidos, entre as quais se destacam funções como oxidação de álcoois, destoxificação de aldeídos/álcoois e metabolismo de ácidos biliares. Uma das subfamílias é a polyol dehydrogenase (PDH), subfamília à qual pertence a enzima XDH e também as sorbitol desidrogenases (SDH; EC 1.1.1.14) e as L-arabinitol 4-desidrogenases provenientes de vários organismos, sendo estas enzimas dependentes exclusivamente do cofator NAD(H) [108, 109].

1.3.6 Análises estruturais às enzimas da via XR/XDH

Existem vários estudos estruturais que já foram feitos a nível das enzimas pertencentes à superfamília AKR, os quais se focavam sobretudo nas aldoses redutases, mais propriamente na compreensão dos seus mecanismos e na sua inibição. Wang et al (2007), construíram um modelo 3D da XR da *S. stipitis* (Código de acesso na UniProt P31867), ou seja, usaram esta sequência de 318 aminoácidos como target, visto que é uma das enzimas responsáveis pela conversão de xilose em etanol e que apresenta um problema que ainda não foi resolvido, que é a acumulação do seu produto final, o xilitol. Aliado a este fator, esta enzima ainda não tem uma estrutura definida experimentalmente, sendo que o uso de ferramentas bioinformáticas para construir um modelo tridimensional para esta proteína se torna a melhor solução,

podendo assim ser efetuados por exemplo estudos sobre os mecanismos de ligação desta enzima e atividade [110].

O template escolhido foi a proteína já cristalizada da XR pertencente ao organismo *Candida tenuis*, a AKR2B5, que pertence à mesma superfamília [110]. A estrutura 3D desta proteína apresenta o código da PDB 1YE4 e contém 322 aminoácidos, um peso molecular de 36 kDa, e 76% de identidade em relação à XR do *S. stipitis* [110,111,112]. No local ativo desta enzima são encontrados os resíduos Tyr-51, Lys-80, His-113 e Asp-46 [112].

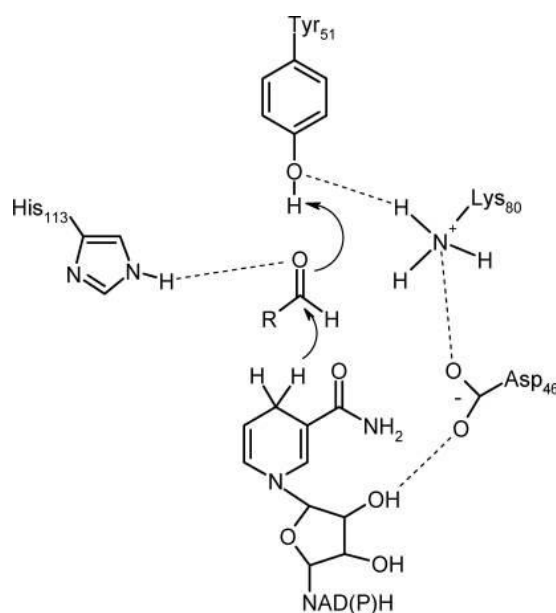


Figura 9 - Mecanismos de reação da xilose redutase da *Candida tenuis* e os resíduos envolvidos na reação [112].

2. MATERIAIS E MÉTODOS

Os métodos usados neste trabalho foram descritos no tópico anterior, todos eles tendo por base conceitos da biologia estrutural e computacional (figura 10).

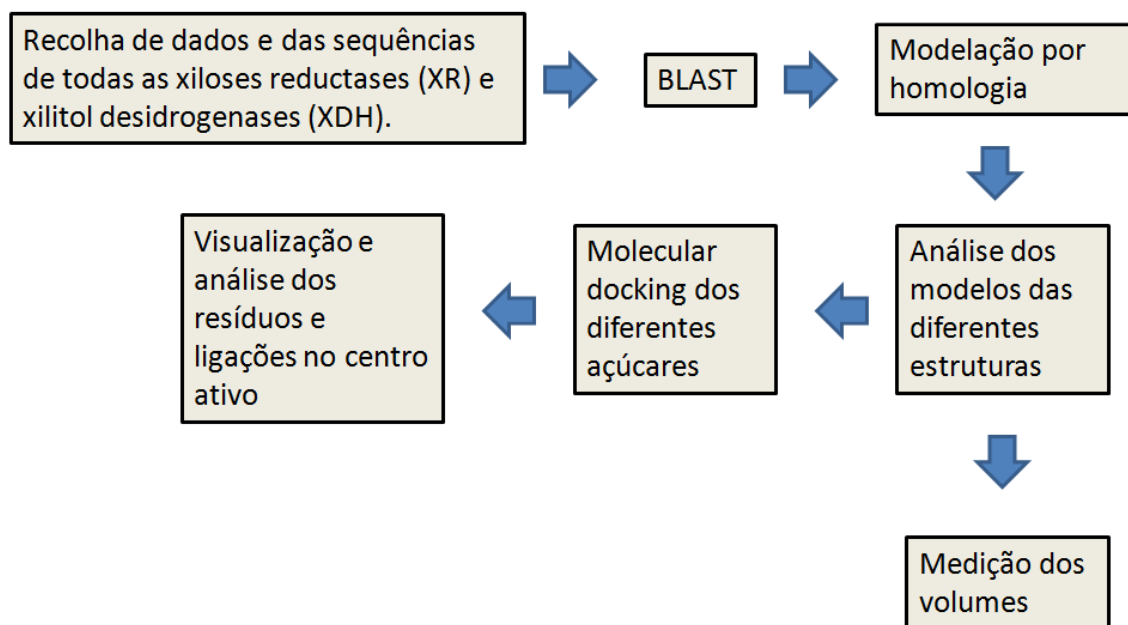


Figura 10 - Esquema e sequência temporal das metodologias usadas para a elaboração deste trabalho.

2.1 Recolha de dados e informação

Para a obtenção de informações sobre as várias xilose reductases (XR) utilizou-se a base de dados Brenda [113] através do número EC 1.1.1.307. Agrupou-se a informação obtida numa tabela com as características dessas várias enzimas que seriam úteis para o resto do trabalho. Esses dados presentes na tabela consistiam: na sequência de aminoácidos, em que nos casos em que ocorreram mutações, estas foram inseridas manualmente nas respetivas posições; no organismo de onde se encontra a proteína; os seus K_M para os diferentes substratos; o código de acesso na base de dados da UniProt [114] e o pH correspondente ao K_M recolhido. Para a análise das similaridade entre as várias sequências foi utilizada a ferramenta clustal ómega [115]. Estes dados foram confirmados nos diferentes artigos e na UniProt. Para alargar a quantidade de enzimas recolhidas recorreu-se à UniProt onde se adicionou enzimas que já foram revistas e anotadas manualmente na base de dados da



SWISS-PROT e que tinham registo dos K_M para alguns substratos. O mesmo procedimento foi efetuado para as xilitol desidrogenases (XDH) onde o número EC utilizado foi o 1.1.1.9.

2.2 Construção dos modelos das enzimas

Todas as sequências previamente encontradas e cuja estrutura ainda não foi decifrada experimentalmente foram submetidas a um Blastp no NCBI comparando com as sequências da base de dados PDB [117]. A PDB contém informação apenas de proteínas cristalizadas, por isso escolheu-se esta base de dados para a obtenção de uma estrutura 3D que apresentasse no mínimo 40% de similaridade com a sequência submetida. O programa usado para gerar os modelos das várias XR e XDH foi o Modeller que requer uma estrutura cristalizada (template), resultado do Blast, e uma sequência duma proteína (target) em formato “ali” [50]. O critério de seleção da estrutura a ser usada como template foi a percentagem maior de similaridade com a sequência, ou seja, os melhores hits do Blast. Além da percentagem de similaridade, o segundo critério obrigatório de seleção da enzima foi esta ter o substrato incluído na sua estrutura. Como nenhuma tinha a xilose ou outro açúcar, e no caso da xilose desidrogenase o xilitol ou outro álcool, usou-se uma enzima que tinha o NAD(P) na sua estrutura cadeia, pois é no carbono 4 do anel de nicotinamida que se dá a reação de protonação [118]. Selecionou-se a cadeia A da “Crystal Structure Of An Engineered K274RN276D Double Mutant of Xylose Reductase From *Candida tenuis* Optimized To Utilize NAD” (1SM9) para todas as XR e também a cadeia A da “Human SDH/NADH inhibitor complex” (1PL6) para todas as XDH. Esta foi escolhida por ser uma sorbitol desidrogenase e ter a mesma função que uma xilitol desidrogenase, sendo mais específica para o sorbitol do que para o xilitol. Além disso, dentro das que possuíam similaridades aceitáveis, era aquela que tinha o NADP na sua estrutura cristalizada.

Antes de proceder à construção do modelo por homologia, era necessário obter um alinhamento entre a estrutura e a sequência. Através de um script no Modeller gerou-se um ficheiro “pap” com esse alinhamento, ficheiro que foi editado para incluir o heteroátomo NAD. Construiu-se de 5 modelos diferentes para cada enzima, dos quais se selecionou aquele com um DOPE score menor.



2.3 Medição dos volumes dos centros ativos das enzimas

Com todos os modelos já construídos procedeu-se a sua análise. O primeiro parâmetro a estudar foi a variação dos volumes dos locais ativos das diferentes enzimas. Com esse intuito utilizou-se o programa *trj_cavity* [52]. Os parâmetros utilizados foram todas as combinações da dimensão (*dim*) entre 6 e 4 juntamente com o *cutoff* a variar entre 6 e 3. Os valores seleccionados foram aqueles que, através da visualização 3D dos ficheiros gerados pelo *trj_cavity* de todas as combinações entre a *dim* e o *cutoff*, se ajustavam melhor ao pocket. Quanto ao *spacing* manteve-se o valor 1.4. Para seleccionar qual dos pockets da enzima que se iria medir definiu-se o *seed* como as coordenadas do C4 do NAD, pois este já estava embutido na estrutura da proteína.

2.4 Docking

O processo de docking dos vários substratos das XR e XDH foi realizado utilizando o programa AutoDock Vina. O processo de docking molecular através do autodock Vina requer os nossos modelos e ligandos em formato “pdbqt” e um ficheiro de configuração “txt”, ao qual foi dado o nome de “conf.txt”. As proteínas foram convertidas em formato “pdbqt”, isto porque em formato “pdb” faltam características como os hidrogénios polares. As proteínas foram preparadas usando as AutoDock tools [119], onde foram adicionados os hidrogénios polares, delimitado o volume onde o substrato iria ser inserido, através da grid box, sendo o seu tamanho foi aproximadamente o dobro das dimensões do ligando (22 Å x 24 Å x 26 Å) no caso da D-xilose e (28 Å x 28 Å x 18 Å) no xilitol. Usando a autodock tool “grid”, as XR e XDH foram convertidas em formato pdbqt. Os vários ficheiros com os ligandos foram conseguidos através da base de dados PubChem onde foram extraídos em formato sdf. Estes ficheiros continham a informação sobre a conformação 3D do substrato. Para poderem ser lidos no Autodock tools foram convertidos em formato PDB. A conversão nesse formato foi realizada através da ferramenta Open Babel [120]. Na ferramenta P.M.V-1.5.6 editou-se o ligando, foram vistas as várias “routable bounds” e os seus diferentes ângulos possíveis de torção e converte-se em formato “pdbqt”. No ficheiro “config.txt” definiu-se como ponto inicial de procura as coordenadas do C4 do NAD, o nome do ficheiro da proteína e do ligando. Este processo foi elaborado para cada uma das xilose reductases e xilitol desidrogenases, em que para cada enzima foi realizado o docking para todos os substratos dos quais disponhamos a informação do K_M . Os scores resultantes do docking foram guardados



num ficheiro “log.txt” e a visualização das conformações dos substratos nas várias proteínas foi feita no PyMol.

2.5 Visualização dos aminoácidos e interações no centro ativo

De modo a visualizar os centros ativos e os substratos usamos o PyMol [121]. Através do comando “find -> polar contacts -> to any atoms” foi possível visualizar as ligações por pontes de hidrogénio é que ocorrem entre o substrato e as cadeias de aminoácidos adjacentes.



3. RESULTADOS E DISCUSSÃO

3.1 Dados da xilose redutase (XR)

Através da pesquisa efetuada foram registados 19 diferentes organismos que possuíam a enzima xilose redutase, dos quais 7 foram rejeitados. As XR dos organismos *Candida intermedia* e *Neurospora crassa*, apesar dos dados recolhidos sobre os seus K_M , não foi possível ter acesso à sua sequência pela base de dados UniProt, o que inabilitou a reconstrução por homologia dos seus modelos estruturais. Relativamente às XR dos organismos *Pachysolen tannophilus*, *Kluyveromyces lactis*, *Emericella nidulans*, *Magnaporthe oryzae* e *Aspergillus oryzae* foram rejeitadas pois não se obteve informação dos seus K_M para a xilose ou outros açúcares.

Dos 11 organismos que foram usados para analisar as características das suas XR, para 3 deles, obteve-se informação pela base de dados BRENDA, de que foram realizadas mutações em certos resíduos para mudar a especificidade do cofator usado pelas XR. A levedura *P.stipitis* (*Scheffersomyces stipitis*) possui 4 diferentes mutantes, sendo que o número de mutações também é variável (Figura 11). Os restantes organismos para os quais também se obtiveram dados relativamente aos seus mutantes foram o *Rasamsonia emersonii* e a *Candida tenuis*, onde apenas há registo de uma mutação. Serão assim usadas um total de 15 sequências de XR neste trabalho.

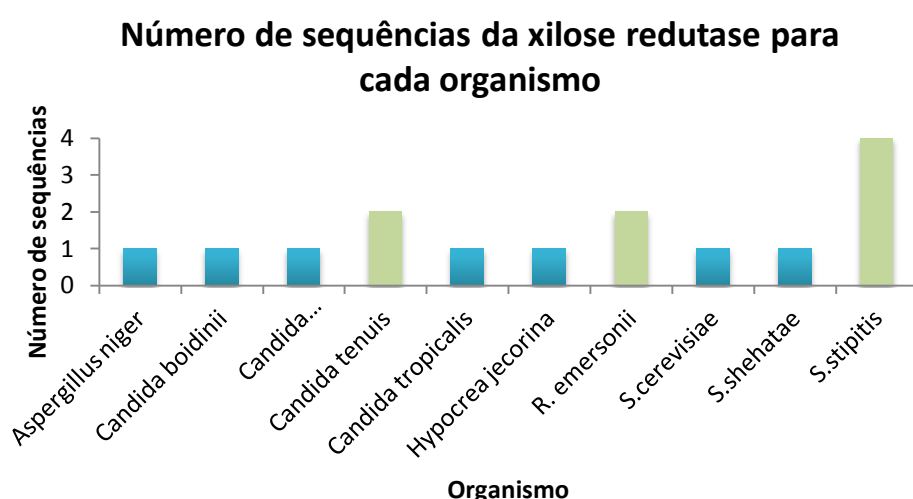


Figura 11- Gráfico com os diferentes organismos e o número respetivo de sequências das xilose redutases obtidas. As barras coloridas a verde indicam aqueles organismos cujas XR possuem mutações, por exemplo, na *Candida tenuis*, foram encontradas duas sequências, a da sua xilose redutase e uma mutação dessa enzima.

Um aspeto que podemos verificar é o facto de mutações que mudam a especificidade do cofator também alterem significativamente o K_M dos açúcares como a xilose. O K_M pode ser definido como a concentração de substrato para que metade dos centros ativos duma amostra dessa enzima estejam ligados ao substrato [122]. No organismo *Rasamsonia emersonii* o K_M da xilose era de 24,6 mM usando o cofator NADP(H) a uma temperatura de 37°C e com um pH de 6,5. Sob as mesmas condições, o K_M do mutante K271R/N273D (o resíduo K e N foram substituídos pelo resíduos R e D nas posições 271 e 273, respetivamente) praticamente triplicou para 76,5 mM. Tais variações também foram observadas para a levedura *Scheffersomyces.stipitis*. Relativamente à *Candida tenuis*, os dados referentes ao K_M da xilose da mutante estão em falta pois não foram encontrados no resultado da pesquisa (figura 12).

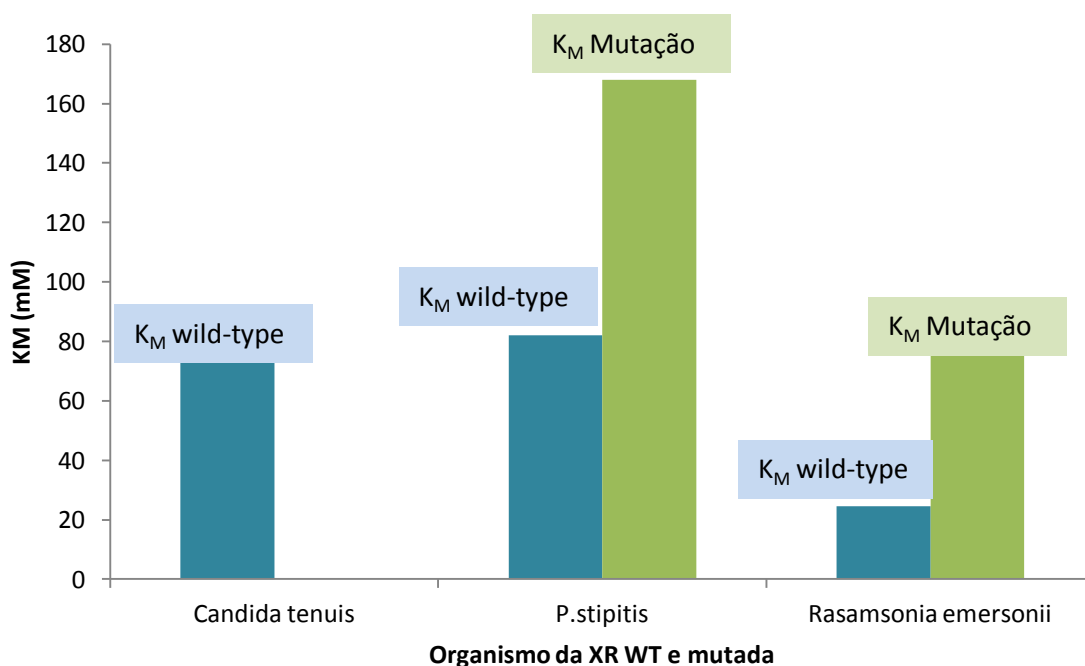


Figura 12 - Gráfico dos K_M em mM da enzima xilose redutase WT (azul) e mutada (verde) nos diferentes organismos. A mutação escolhida para a *P.stipitis* (*Scheffersomyces stipitis*) foi K270S/N272P/S271G/R276F e para o organismo *Rasamsonia emersonii* foi a K271R/N273D.

Em algumas XR dos diferentes organismos foi obtida a informação sobre os K_M de vários açúcares que estas utilizam como substrato, onde posteriormente iriam ser sujeitos ao processo de *docking* molecular para serem comparadas características como afinidades, modos de ligação e tamanho.

Para todos os organismos que foram encontrados na Brenda conseguimos obter as condições físico-químicas a que estão sujeitas pois estas características têm grande influência



nas afinidades e por sua vez nos K_M , tais como o pH, o co-substrato utilizado e a temperatura. Estes são diferentes pois o pH e temperatura ótima variam entre os diversos organismos e enzimas [122,123]. Isto é importante pois eventuais comparações de K_M entre as várias enzimas pode ser um fator redundante. Os dados obtidos em relação a estas características físico-químicas não significam que sejam sempre os ótimos, podendo levar a diferenças entre os K_M .

As sequências encontradas contêm tamanhos semelhantes, compreendidos entre 318 e 329 aminoácidos (Figura 13A). Foi possível observar, recorrendo posteriormente a métodos de docking que, possivelmente, certos aminoácidos que interagem com os substratos no centro ativo estão conservados em todas as XR, evidenciando mesmos mecanismos catalíticos. Já através da análise da árvore filogenética, as duas sequências mais semelhantes são as da xilose redutase da levedura *Scheffersomyces shehatae* e da *Candida tenuis*, com tamanhos de 323 e 322, respectivamente (Figura 14B).



A

Aspergillus_niger_Q9P8R5_XR	----MASPTVKLNSGYDHPVLVGFGLKKNNDTCADQIYHAIKEGYRLFQDGAQYVNEVEA	56
C533R6_Rasamsonia_emersonii_XR	----MATPTIKLNSGYDHPVLVGFGLKKNRETCAQVYEAIKAGYRLFQDGAQYVNEVEA	56
P38715 Saccharomyces_cerevisiae_XR	----MSSLVTLNNGKLNPLVGLGCKNKIDKKVCANQIYEAIKLGYRLFQDGAQYVNEKEV	55
P49378_Kluyveromyces_lactis_XR	--MTYLAETVTLNNGKLNPLVGLGCKNMPNDVCAQIYEAIKIGYRLFQDGAQYVNEKEV	58
P78736_Pachysolen_tannophilus_XR	----MTLQVYTLNNGRIKIPAIKIGMCKLEN--AADMVYAAIKEGYRLFQDGAQYVNEKEV	54
Q8X195_Candida_boidinii_XR	----MSSPLLTLNNGKLNPKQIGFGCKNKVDNATCAETIYEAIKVGYRLFQDGAQYVNEKEV	56
Q6Y0Z3_Candida_parapsilosis_XR	MSTATASPAVKLNSGYEIPVLVGFGLKLTNDVASDQIYRAIKSGYRLFQDGAQYVNEQEV	60
P31867_Scheffersomyces_stipitidis_XR	-----MPSIKLNSGYDHPAVGFGCKNKVDVTCSEQIYRAIKTGVALFDGAQYVNEKLV	54
P87039_Candida_tropicalis_XR	MSTTPTIPTIKLNSGYENPLVGFGLKKNVNTAADQIYNNAIKTGVALFDGAQYVNEKEV	60
Q9P430_Scheffersomyces_shehatae_XR	-MSPSPIPAFKLNNGLENPSIGFGCKLKDSTAADQVYNAIKAGYRLFQDGAQYVNEQEV	59
O74237_Candida_tenuis_XR	--MSASIPDKLSSGHLNPSIGFGCKLANATAGEQVYQAIKAGYRLFQDGAQYVNEKEV	58
	..*.* :* :*.* ** .. :* ** ***** * **	
Q876L8_Hypocrea_jeconina_XR	GEGVARAIKDGVLKREDFIVSKLWQTFHDEQKVEPIRRQLADWQIDYDFDLFLHPIA	116
Aspergillus_niger_Q9P8R5_XR	GQGIARAIKDGVLKREELFIVSKLWNSFHDGDRVEPICRQLADWGIQDYDFDLYHPIA	116
C533R6_Rasamsonia_emersonii_XR	GQGVARAIKEGIVKREDFIVSKLWNTFHEADKVEPIARQLADWGLDYDFDLYHPIA	116
P38715 Saccharomyces_cerevisiae_XR	GEGIRKAIKSEGLVSRKIDFVSVKLWNNFHPDHVKLALKKTLSDMGLDYDFDLYHPIA	115
P49378_Kluyveromyces_lactis_XR	GQGVNRAIKEGVLKREDFLVVSVKLWNSFHHPDNVPRALERTLSDLQLDYVDIFYHPIA	118
P78736_Pachysolen_tannophilus_XR	GEGINRAIKDGVLKRDLFITSKLWNNFHAKEVVKKALKMSLSDFNLDYDFDLYHPIA	114
Q8X195_Candida_boidinii_XR	GEGVNIAIKDGVLKREELFIVSKLWNNFHPDSVKLAIKKVLSDLNLDYDFDLYHPIA	116
Q6Y0Z3_Candida_parapsilosis_XR	GEGIKRAIKEGIVKREELFIVSKLWNSFHDKNVEVALNKTLSDNLDYVDIFYHPIA	120
P31867_Scheffersomyces_stipitidis_XR	GAGVKKAIKDEGIVKREDFLTSKLWNNYHHPDNVEKALNRTLSDLQVDYVDIFYHPIA	114
P87039_Candida_tropicalis_XR	GEGINRAIKEGVLKREELFIVSKLWNNFHPDNVETALNKTLSDNLDYVDIFYHPIA	120
Q9P430_Scheffersomyces_shehatae_XR	GEGVKRAIKDEGIVTRREEIFITSKLWNNYHHPDNVETALNKTLKDLKVDYVDIFYHPIA	119
O74237_Candida_tenuis_XR	GQGVKRAIKDEGLVKREELFIVSKLWNNYHHPDNVETALNKTLADLKVVDYVDIFYHPIA	118
	. :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.*	
Q876L8_Hypocrea_jeconina_XR	LEYVDPSVRYPGPFYDGKSE----VRWSKTTTLQQTWAGMERLVOKGLARSIGVSNYQA	172
Aspergillus_niger_Q9P8R5_XR	LKYVDPAVRYPGPFYDGKSEK-DE----LEFG-NATIQETWAMESLVOKGLARSIGISNFS	170
C533R6_Rasamsonia_emersonii_XR	LKYVDPAEYPPGWTGK-KE----VEFS-NATIQETWQAMETLVOKGLARSIGISNFS	170
P38715 Saccharomyces_cerevisiae_XR	FKYVPFEKYPGPFYTGADDEKKGHITA-HVPIIDTYRALEECVDEGLIKSIGVSNFQG	174
P49378_Kluyveromyces_lactis_XR	FKPVPFDEKYPGPFYTGKDEAKGHIEE-QVPLLDTWRALEKLVQDQGLIKSIGISNFSG	177
P78736_Pachysolen_tannophilus_XR	FKFVPFEKYPGPFYCGDGDK----FIYE-DVPIIETWRAMENLVDEGLVKSIGVSNVSG	169
Q8X195_Candida_boidinii_XR	QKFVPTEKKYPGPFYCGDGDK----WSFE-DVPLLTWRAMEELVEGLVKSIGISNFGV	171
Q6Y0Z3_Candida_parapsilosis_XR	QKVPVTEKKYPGPFYCGDGDK----WSIE-EVPLLDTWRALEKLVQDQGLIKSIGISNFS	175
P31867_Scheffersomyces_stipitidis_XR	FKFVPLEEKYPGPFYCGDGDN----FDYE-DVPILETWKALEKLVKAGIKSIGVSNFPG	169
P87039_Candida_tropicalis_XR	FKFVPTEKKYPGPFYCGDGDN----FHYE-DVPLLDTWRALEKLVKAGIKSIGISNFTG	175
Q9P430_Scheffersomyces_shehatae_XR	FKFVPIEEKYPGPFYCGDGDN----FVYE-DVPILETWKALEKLVKAGIKSIGVSNFPG	174
O74237_Candida_tenuis_XR	FKFVPIEEKYPGPFYCGDGN----FVYE-DVPILETWKALEKLVKAGIKSIGVSNFPG	173
	:* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.*	
Q876L8_Hypocrea_jeconina_XR	QSVYDALIVARIKPATLQIEHHPYLTQQLVSLAQTEGIVVTAYSSFGPTGFNEMDMPRA	232
Aspergillus_niger_Q9P8R5_XR	QLVNDLLRYARIRPATLQIEHHPYLTQQLVSLAQTEGIVVTAYSSFGPLSFLSLVQNA	230
C533R6_Rasamsonia_emersonii_XR	QLINDLLRYARIRPATLQIEHHPYLTQQLVSLAQTEGIVVTAYSSFGPLSFLSLVQNA	230
P38715 Saccharomyces_cerevisiae_XR	SLIQDQLLRGCRIPKVALQIEHHPYLTQEHLEFCKLHDIQVVAAYSSFGPQSFIEHDLQLA	234
P49378_Kluyveromyces_lactis_XR	ALIQLDLRGARIPKVALQIEHHPYLTQERLIKVKNAQIQVVAAYSSFGPVSFLELENKKA	237
P78736_Pachysolen_tannophilus_XR	GLLEDLIKAARIKPAQLQIEHHPYLTQQLVSLAQTEGIVVTAYSSFGPLSFLSLVQNA	229
Q8X195_Candida_boidinii_XR	ALIQLDLRGARIPKVALQIEHHPYLTQERLIKVKNAQIQVVAAYSSFGPVSFLELENKKA	231
Q6Y0Z3_Candida_parapsilosis_XR	QLIYDLIRGCTIKPVALQIEHHPYLTQPKLVYVQLHDIQITGYSSFGPQSFLEHDLKRA	235
P31867_Scheffersomyces_stipitidis_XR	ALLLDLLRGATIKPSVLQVEHHPYLTQQLRIEFAQSRGIAVTAYSSFGPQSFVFNQGR	229
P87039_Candida_tropicalis_XR	ALIYDLIRGATIKPAVLQIEHHPYLTQPKLIEVYQKAGIAITGYSSFGPQSFLELESKRA	235
Q9P430_Scheffersomyces_shehatae_XR	ALLLDLFRGATIKPAVLQVEHHPYLTQPKLIEYQKAGITVTAYSSFGPQSFVFNQGR	234
O74237_Candida_tenuis_XR	ALLLDLLRGATIKPAVLQVEHHPYLTQPKLIEFAQKAGVTITAYSSFGPQSFVFNQGR	233
	:* :* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.* :*.*	

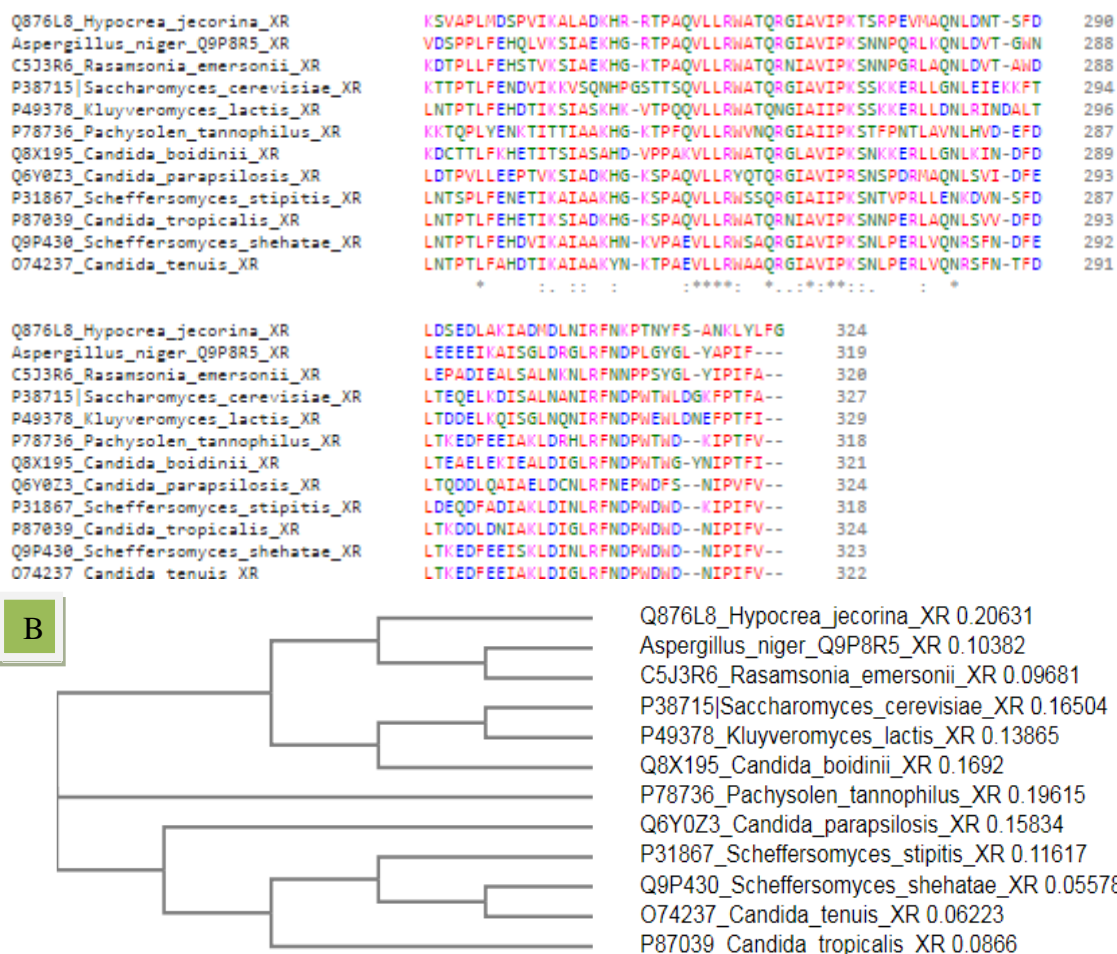


Figura 13 - Alinhamento de todas as sequências das xilose redutases com os seus respectivos tamanhos. Realçado a vermelho aparecem aminoácidos comuns nos centros ativos de todas as enzimas e intervenientes na reação de redução da xilose (A), usando o clustal omega [115]. Árvore filogenética das sequências xilose redutases dos vários organismos e respetivos scores das distâncias (B).

3.2 Estruturas tridimensionais das xilose redutases

Os resultados do Blast da similaridade das sequências recolhidas com as guardadas na PDB indicam que todas as proteínas encontradas com maior identidade servem de template para a construção dos modelos das enzimas, pois todas cumprem o requisito de possuírem uma similaridade superior a 40% [45]. De entre todas as xilose redutases usadas como template apenas uma possuía 100% de similaridade com o target, a xilose redutase da *Candida tenuis*, cujo UniProt id é O74237 e o da PDB 1JEZ (Figura 14).

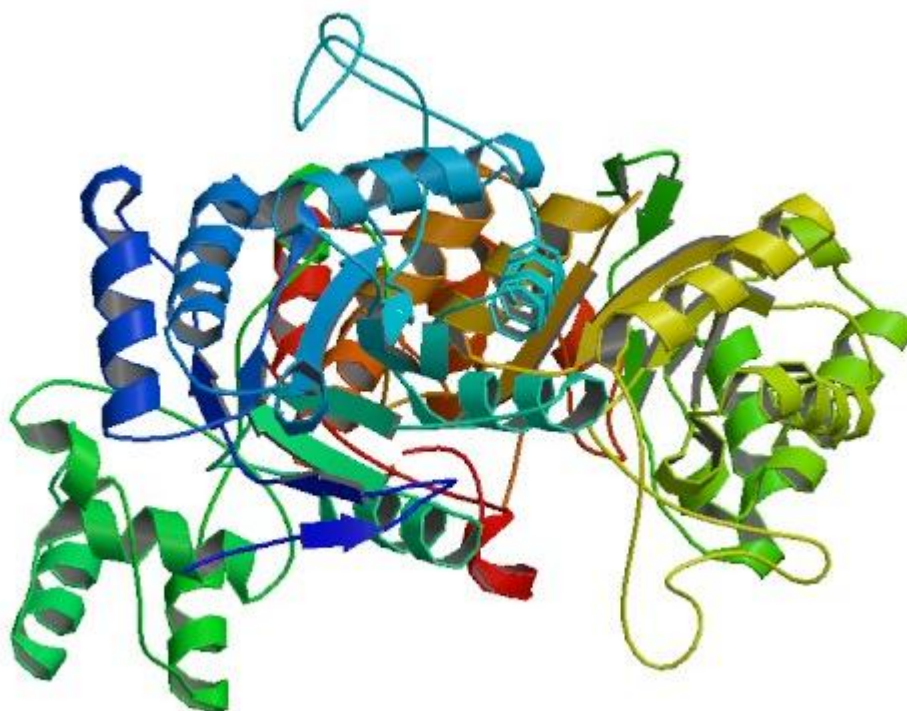


Figura 14 - Estrutura cristalizada da xilose redutase dimérica com o cofator NAD⁺ do organismo *Candida tenuis* (1JEZ) [92].

Através da análise dos modelos criados foi possível verificar que todos eles apresentam o mesmo *fold* (Figura 15).

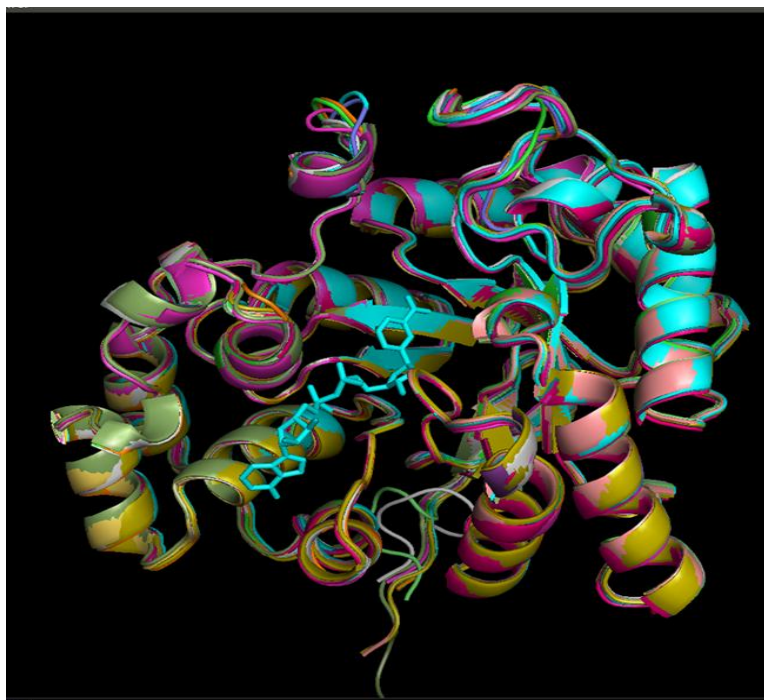


Figura 15- Sobreposição de todas as estruturas da xilose redutase com o NAD representado a azul claro.

3.3 Dados da xilitol desidrogenase

Foram obtidas 17 xilitol desidrogenases (XDH) de diferentes organismos, em que para todos eles foi conseguida informação dos valores seus K_M . Porém, não foi possível encontrar a sequência de todos devido à falta do UniProt id. Por estes motivos, os resultados da pesquisa permitiram a obtenção de 7 XDH diferentes, das quais duas foram conseguidos dados sobre algumas mutações que alteravam a afinidade para o xilitol e outros substratos. Essas mutações visavam alterar o cofator da enzima [124,125,126]. Para *Scheffersomyces stipitis* foram encontrados uma vasta quantidade de mutantes sendo que para o organismo *Rhizomucor pusillus* essa quantidade foi menor, 13 e 2 respectivamente (Figura 16). Os dados obtidos para outros álcoois não foram tão diversos como para os diferentes substratos das XR, sendo que apenas o ribitol consumido pelos organismos *Serratia marcescens* e *Rhizomucor pusillus* e o sorbitol pelo *Blastobotrys adeninivoran*, *Rhizomucor pusillus* e *Candida sphA* são os únicos que temos dados para os seus K_M para além do xilitol.

XDH mutadas por organismo

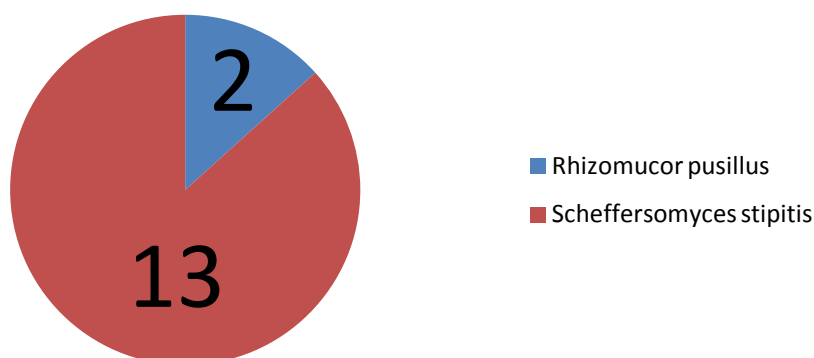


Figura 16 - Gráfico circular do número de enzimas XDH (incluindo a WT) e respectivos organismos a que pertencem.

Tal como os resultados da XR, também na XDH as propriedades físico-químicas diferem de organismo para organismo bem como nos vários mutantes. Em *Scheffersomyces stipitis*, apesar do pH ser de 9 para todos os mutantes a uma temperatura de 35°C, o cofator usado varia entre o NAD^+ e o $NADP^+$, o que influencia significativamente os K_M para a xilose. Como é possível observar no gráfico da Figura 17, para a mesma mutação e sob as mesmas condições físico-químicas mas com co-substratos diferentes, por exemplo, o K_M do xilitol do mutante D207A/I208R/F209T passa de 0,638 mM com o $NADP^+$ para 97,8 mM com o NAD^+ como cofator (Figura 17).

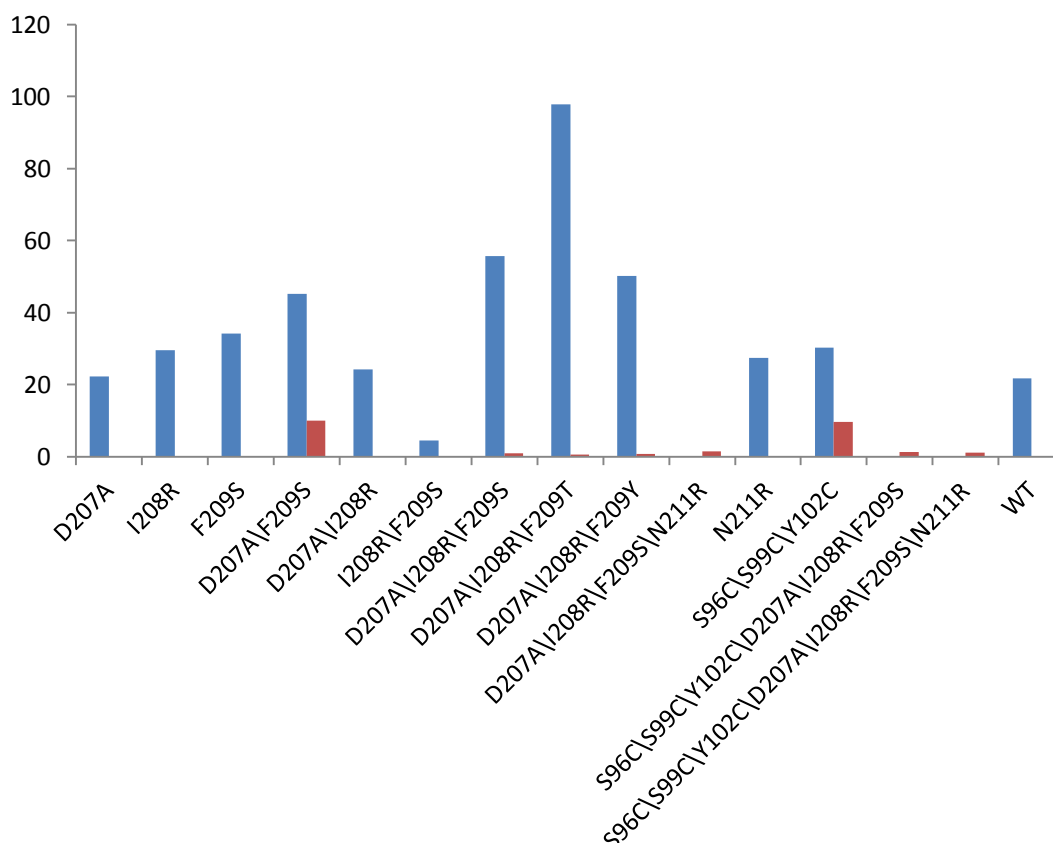


Figura 17 - Gráfico que compara os diferentes K_M da xilose da *Scheffersomyces stipitis* e dos diferentes mutantes quando esta se encontra ligada ao NAD^+ (azul) ou $NADP^+$ (vermelho). Para alguns mutantes foi apenas conseguido encontrar dados para o K_M de um dos cofatores.

Relativamente às sequências recolhidas, estas têm tamanhos entre 344 e 364 resíduos. Há duas enzimas que apresentam o mesmo tamanho, a XDH da levedura *Scheffersomyces stipitis* e do *Rhizomucor pusillus*, ambos com 363 aminoácidos. Porém, através da análise da árvore filogenética, verifica-se que estas não se encontram no mesmo cluster inicial, ou seja, não agrupam na primeira ramificação, evidenciando que há outras sequências mais semelhantes apesar dos seus diferentes comprimentos. A XDH do *Rhizomucor pusillus* a que mais difere de todas as outras (Figura 18).



Figura 18 - Árvore filogenética das xilitol desidrogenases dos diferentes organismos e respectivos scores, usando a ferramenta clustal omega [115].

3.4 Estruturas tridimensionais das xilitol desidrogenases

Através dos resultados do Blast, apenas a xilitol desidrogenase da *Candida tropicalis* possui uma identidade abaixo da requerida para obter um modelo de qualidade. Tal como nas XR foi encontrada uma XDH já cristalizada, pertencente ao organismo *Gluconobacter oxydans* e com o UniProt id Q8GR61 e 1ZEM na PDB (Figura 19).

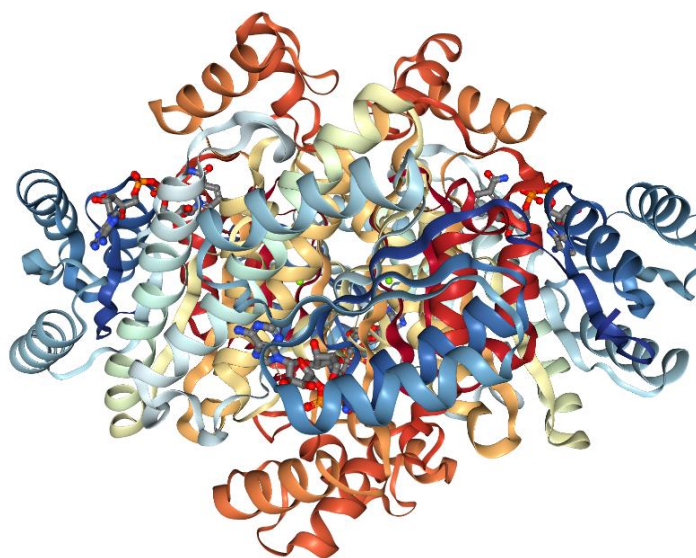


Figura 19 - Estrutura cristalizada da xilitol desidrogenase com o cofator NAD⁺ do organismo *Gluconobacter oxydans* (1ZEM) usando o NGL viewer [127].

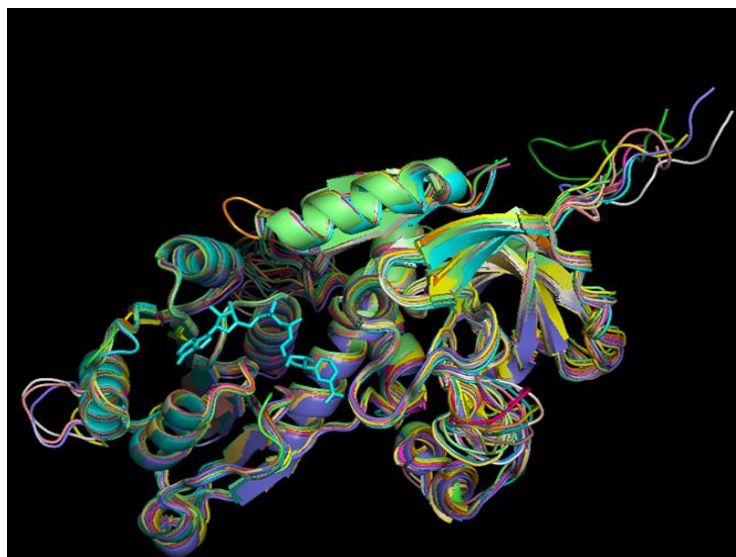


Figura 20 - Sobreposição de todas as estruturas secundárias da xilitol desidrogenases com o NAD representado a azul claro. Em ambas as figuras é possível visualizar o *fold* das proteínas.

3.5 Volumes dos centros ativos das xilose redutases e xilitol desidrogenases

Os volumes dos locais ativos variam, apesar de todas as enzimas serem homólogas entre si e possuírem os mesmos *fold* (Figuras 15 e 20). Os volumes das diversas xilose redutases apresentam uma gama de resultados entre 110 Å³ e 266 Å³, sendo a XR da *Candida tenuis* aquela com um menor volume e a do *Aspergillus niger* a que têm um volume de maiores dimensões (Tabela 2).

Tabela 2 - Tabela dos organismos a que pertencem as várias xilose redutases e valor dos respetivos volumes medidos pelo trj_cavity.

Organismo	Volume em Å ³
<i>Aspergillus niger</i>	266,168
<i>Candida boidinii</i>	150,920
<i>Candida parapsilosis</i>	186,592
<i>Candida tenuis</i>	109,760
<i>Candida tenuis</i> , mutant enzyme D50A	142,688
<i>Candida tropicalis</i>	192,080
<i>Hypocrea jecorina</i>	200,312
<i>Rasamsonia emersonii</i>	194,824
<i>Rasamsonia emersonii</i> , mutant K271R/N273	170,128
<i>Saccharomyces cerevisiae</i>	194,824
<i>Scheffersomyces shehatae</i>	159,152
<i>Scheffersomyces stipitis</i>	123,480
<i>Scheffersomyces stipitis</i> , mutante K21A/N272D	167,384
<i>Scheffersomyces stipitis</i> , mutante K21A	148,176
<i>Scheffersomyces stipitis</i> , mutante K270S/N272P/S271G/R276F	132,000



Como todas as proteínas são homólogas ou mutações das mesmas, através da sua sobreposição é possível verificar que as diferenças estruturais não são muitas apesar de existirem, apresentando todas o mesmo *fold*, pelo que a utilização do parâmetro dim 5 e cutoff 5 foi a que demonstrou ter uma medição de volume dos centros ativos mais correta para todas as XR. Quando o algoritmo do programa trj_cavity é aplicado várias vezes para a mesma proteína com os mesmos parâmetros, o ficheiro resultante apresenta sempre o mesmo volume. Isto sugere que o algoritmo de procura do trj_cavity deteta ligeiras mudanças conformacionais entre as várias proteínas e por sua vez diferentes volumes, pois apenas a xilose redutase da *Saccharomyces cerevisiae* e do organismo *Rasamsonia emersonii*, apresentam o mesmo volume de 195 Å³ (Tabela 2). Apesar desta igualdade, as duas enzimas na árvore filogenética não eram as que apresentavam a sequência mais semelhante entre si.

Os volumes do organismo *Aspergillus niger* e do mutante K271R/N273D da *Rasamsonia emersonii*, não se apresentam corretamente medidos como é possível observar na figura (figura 21). As suas cavidades transpõem a estrutura medindo partes da estrutura que já não pertencem ao local ativo. Estes resultados podem dever-se ao facto do programa usado, o trj_cavity, não ter conseguido delimitar corretamente o centro ativo, apesar de terem sido testados todos os parâmetros, tal como foi feito para todas as outras xilose redutases. Podemos ainda salientar que, para proteína não mutante do organismo *R.emesonii*, conseguimos determinar com uma melhor eficácia o volume, ou seja, a dupla mutação K271R/N273D têm uma influência direta no volume desta proteína.

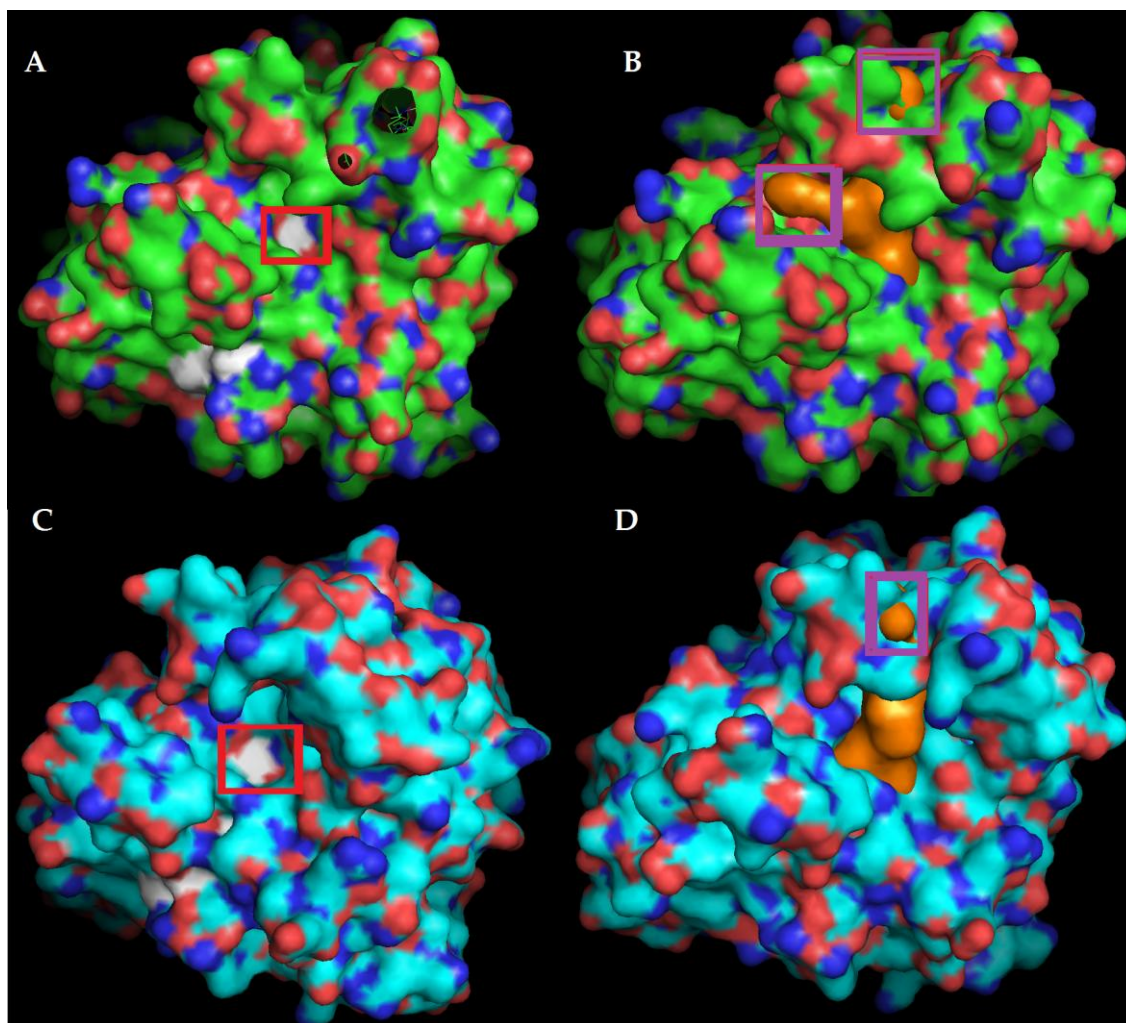


Figura 21 - Estruturas em superfície da xilose redutase do *Aspergillus niger* (A, B) e do mutante K271R/N273D da *Rasamsonia emersonii* (C, D), ambas com os respectivos volumes dos centros ativos representados a laranja. Nas imagens A e C o NADH está embutido na estrutura a branco e está delimitado a vermelho o local onde se encontra o C4 do NADH, ao qual se vai ligar a xilose. Nas imagens B e D encontra-se delimitado a roxo o local onde o volume foi mal medido e transpõe o centro ativo.

Uma das possíveis incongruências na medição dos volumes foi não se ter em consideração as dinâmicas moleculares das moléculas, podendo ocorrer possíveis modificações estruturais nas proteínas, quer na ligação do cofator, quer na ligação da xilose ou outro açúcar. Para isso seria necessário obter informação estrutural sobre o comportamento da proteína com esses substratos ou tentar prever eventuais mudanças estruturais através das ligações que esta enzima estabelece com os vários substratos. Assim, prevendo a trajetória ao longo do tempo do substrato na proteína através dum programa computacional correto seria ideal para este problema e minimizaria eventuais erros de medição bem como aumentaria a precisão. O programa *trj_cavity* possui um parâmetro “f” que corresponde a um ficheiro com as trajetórias moleculares que daria um output do volume da cavidade da proteína ao longo do tempo [52]. Poderiam também ter sido usados outros programas que usam MD para



identificar os volumes dos centros ativos como o Caver3.0, POVME [128], Epock, dxTuber, MDpocket.

Pela análise do gráfico da Figura 22 é possível observar uma tendência para uma diminuição do volume à medida que o K_M aumenta (Figura 22), indicado pelo declive de -1,0625. Tal poderá dever-se ao facto de a enzima ao ter um maior volume, o substrato possuir mais espaço para arranjar uma conformação mais estável. Porém, o R^2 é 0,3709, o que não permite estabelecer uma forte relação de que o volume influencia diretamente o K_M . O local ativo das enzimas é um local com propriedades físico-químicas muito particulares, e os registos dos valores experimentais variam em certos parâmetros. Um exemplo é o caso dos mutantes da *Scheffersomyces stipitis*, para o mutante K270S/N272P/S271G/R276F o pH é 6,0 enquanto o dos outros 2 mutantes e WT é 6,5 bem como o cofator preferencial. Isto aliado a erros no programa por transpor a superfície das proteínas e também à não consideração das dinâmicas moleculares ao longo do tempo podem ser os fatores que interferiram com os resultados obtidos.

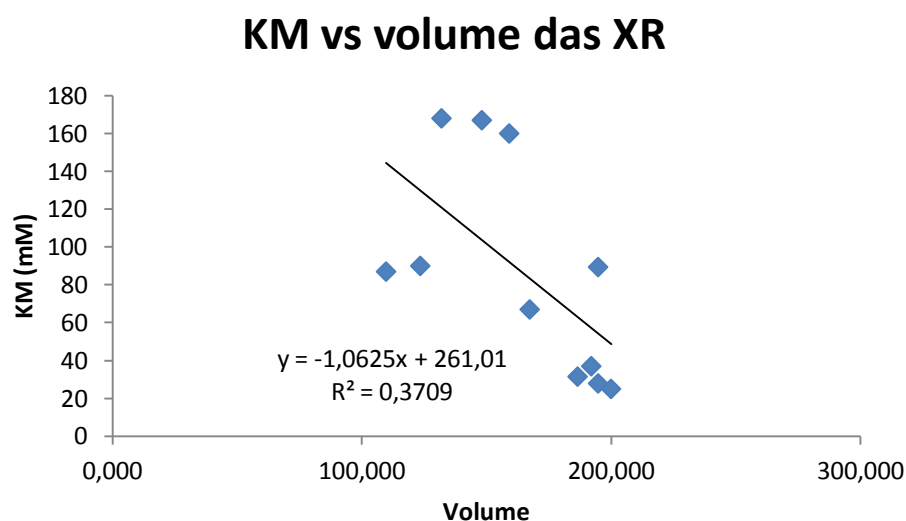


Figura 22 - Gráfico da influência do volume das XR no seu K_M . Cada ponto azul representa o volume e respetivo K_M para cada uma das xilose redutases dos diferentes organismos e também os seus mutantes. Os volumes que foram mal delimitados foram descartados deste gráfico, obtendo a equação $y = -1,0625x + 261,01$ com um R^2 de 0,3709.

3.6 Volume dos vários mutantes da levedura *Scheffersomyces stipitis*

A xilose redutase tem preferência para a utilização do NADPH como cofator em relação ao NADH [111]. No mutante K21A da levedura *Scheffersomyces stipitis*, a lisina foi mutada por uma alanina. No local ativo da XR, o resíduo Lisina, na posição 21 da sequência, estabelece fortes ligações por pontes de hidrogénio com o NADPH, contrariamente ao que se



sucede com o NADH [124]. Assim, esta mutação é eficaz para alterar o cofator preferencial para o NADH. O mesmo acontece com o duplo mutante K21A/N272D, no qual esta alteração de cofator ainda se demonstrou mais eficiente. Como as dimensões e volumes do NADPH e NADH são praticamente as mesmas, era de esperar que o volume dos mutantes e da enzima não mutada fossem todos muito semelhantes, o que não se sucedeu, pois estes variam de 124 Å³ a 167 Å³. Tais diferenças podem advir das diferentes dimensões das cadeias laterais dos aminoácidos que foram mutados. É expectável que aminoácidos mais pequenos a envolver os locais ativos lhes confiram um maior volume do que os aminoácidos de maior tamanho, pois estes últimos vão ocupar um maior volume. A lisina apresenta um tamanho e peso molecular maior do que alanina, 146,19 g/mol e 89,094 g/mol respetivamente. Estes resultados vão de encontro ao que era previsto, pois o volume da XR não mutada da *S.stipitis* é de 124 Å³ e a mutante K21A é de 148 Å³, sendo este possivelmente um dos fatores que influencia o volume do local ativo. No mutante K21A e N272D, como já foi referido anteriormente, a lisina é bastante maior que a alanina, mas na mutação que foi acrescentada, a troca de uma asparagina pelo ácido aspártico na posição 272, não deveria influenciar o volume significativamente pois estes dois aminoácidos possuem praticamente as mesmas dimensões e pesos moleculares. Outro fator que pode influenciar estes resultados é que a troca de resíduos muda a conformação do centro ativo na modelação, como por exemplo as cadeias laterais ficarem virados mais para o interior ou exterior da enzima. Tal não se verificou pois o volume com esta mutação passou a ser de 167 Å³, o que provavelmente evidencia a presença de outros fatores que terão influência no volume do centro ativo, tal como discutido na secção anterior. Juntamente a estes fatores, e analisando o gráfico da Figura 23 é possível verificar que não foi conseguida estabelecer uma relação entre os volumes dos diferentes mutantes e o K_M (Figura 23), pois o R^2 é apenas de 0,1063.

Volume dos mutantes da P.Stipitis da XR vs Km

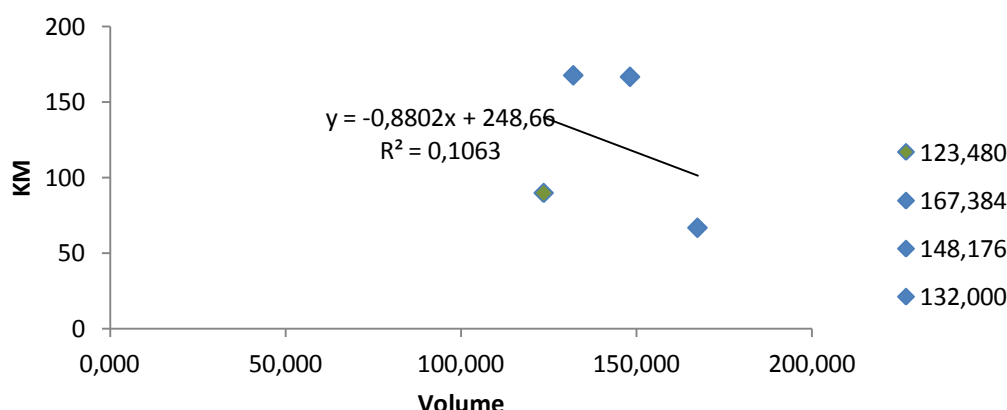


Figura 23 - Gráfico da influência do volume das XR no seu K_M para os mutantes da levedura *Scheffersomyces stipitis*. Cada ponto azul representa o volume e respetivo K_M para cada uma das 3 xilose redutases mutadas deste organismo. O ponto verde corresponde à enzima não mutada. Este gráfico apresenta a equação $y = -0,8802x + 248,66$ com um R^2 de 0,1063.

3.7 Volumes dos centros ativos das xilitol desidrogenases

Relativamente à XDH, os volumes do seu centro ativo variam entre 28 \AA^3 do organismo *Candida sp. HA* e 118 \AA^3 do triplo mutante D207A/I208R/F209Y da levedura *Scheffersomyces stipitis*. Dentro das 13 mutações da *S.stipitis* apenas duas continuaram com o mesmo volume da levedura WT, o mutante S96C/S99C/Y102C/D207A/I208R/F209S/N211R e o D207A/I208R/F209S/N211R com um volume de 102 \AA^3 . Pode-se inferir que a mutação D207A provavelmente não tem qualquer influência no volume do centro ativo da enzima onde ocorre a ligação do xilitol pois o volume do mutante I208R e do duplo mutante D207A/I208R é para ambos os casos de 74 \AA^3 .

Apesar dos resultados supracitados, a xilitol desidrogenase tem um centro ativo mais intrínseco e fechado do que a XR, o que pode ter influência na medição dos volumes para estas enzimas ser menos precisa. Apesar de testados todos os parâmetros, em nenhum foi conseguido uma boa delimitação total do centro ativo.

Os resultados da análise dos volumes da XDH não permitem afirmar a possibilidade de o volume do centro ativo ser ou não um fator que influencia as diferentes afinidades para os vários substratos e dos seus K_M . Relativamente aos resultados dos volumes das XR, parecem evidenciar que o volume não é um fator preponderante para as variações do K_M dos diferentes substratos, porém tal não consegue ser fortemente suportado pelos nossos resultados. Isto pode dever-se sobretudo à ineficácia do método usado pelas razões explicadas anteriormente.



Deste modo foi usada uma diferente abordagem para descobrir características dos substratos que influenciam as suas diferentes afinidades na sua ligação à xilitol desidrogenase e também à xilose redutase, o *docking* molecular.

3.8 Docking da xilose nas xilose redutases

Pela análise do score e visual de todas as xilose redutases sujeitas ao *docking* molecular, foi a XR dos organismos *Candida boidinii* e *Candida tenuis* com um score de -5,1 (“binding affinity” em Kcal/mol) o que pode ser um indicador de que são estas enzimas que estabelecem ligações mais fortes e têm uma maior afinidade para a xilose, apesar das diferenças dos valores do score não serem muito acentuadas (Tabela 3).

Tabela 3 - Os scores do docking da D-xilose nas diferentes xilose redutases e respetivos aminoácidos que estabelecem interações polares.

Organismo	Score do Docking	Aminoácidos que estabelecem interações polares no centro ativo
<i>Aspergillus niger</i>	-4,6	His Ala
<i>Candida boidinii</i>	-5,1	Asn His
<i>Candida parapsilosis</i>	-4,9	Asn His
<i>Candida tenuis</i>	-5,1	His Trp
<i>Candida tenuis</i> , mutant enzyme D50A	-5,0	Asn His
<i>Candida tropicalis</i>	-4,8	Asn His Tyr Trp
<i>Hypocrea jecorina</i>	-4,5	Asn His
<i>Rasamsonia emersonii</i>	-4,6	Asn His Tyr
<i>Rasamsonia emersonii</i> , mutant K271R/N273	-4,5	Asn His Tyr
<i>Saccharomyces cerevisiae</i>	-5,0	Asn His Tyr
<i>Scheffersomyces shehatae</i>	-4,6	His Tyr
<i>Scheffersomyces stipitis</i>	-5,0	His Tyr
<i>S.stipitis</i> , mutante K21A/N272D	-4,6	His Asp Tyr
<i>S.stipitis</i> , mutante K21A	-4,8	His
<i>S.stipitis</i> , mutante K270S/N272P/S271G/R276F	-4,8	His Tyr

3.9 A influência do volume do centro ativo nos resultados do *docking*

Pela análise do gráfico da figura 24 é possível observar uma tendência de quanto maior for o volume, menor será o score em módulo, ou seja, pior afinidade para a D-xilose. Estes resultados vão de encontro ao esperado. Se anteriormente foi comparado os volumes aos K_M e verificou-se que, provavelmente, quanto maior fosse o volume maior seriam as opções para se ligar, logo teria um K_M mais baixo (maior afinidade), no caso do *docking* há um fator diferente, a “grid box”. A “grid box” é o espaço que o ligando tem para se mover e testar diferentes conformações. Para todas as XR, o substrato sujeito ao *docking* era o mesmo, a D-xilose. Por este motivo definiu-se o espaço para se mover (“grid-box”) como sendo o mesmo para todas as XR. Esta “grid box” era mais pequena que o total do centro ativo, e por este fator, a D-xilose iria ficar mais afastada de certos resíduos em volumes maiores, levando a scores de *docking* menores (em módulo).

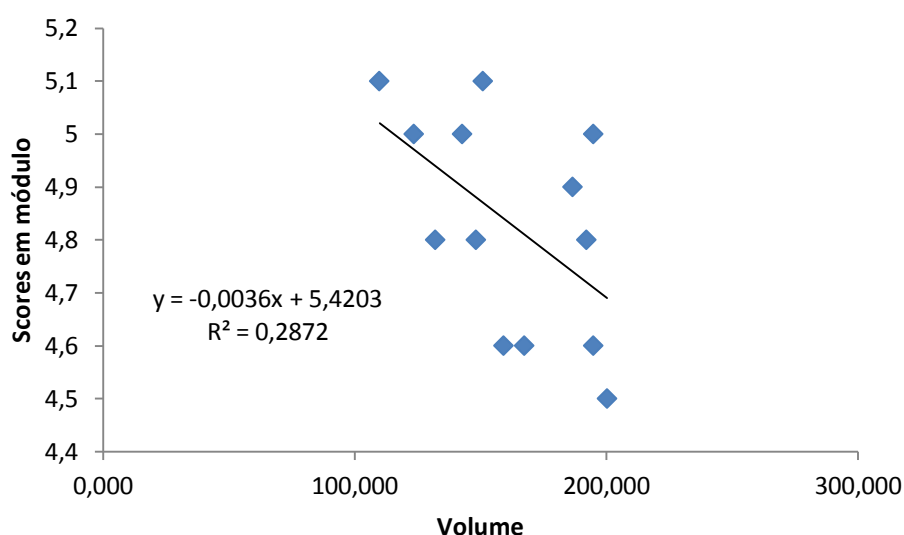


Figura 24 – Regressão linear entre os scores do *docking* e o volume. A equação da reta apresenta um declive negativo, sendo: $-0,0036x + 5,4203$, com um R^2 de 2,872.

3.10 Docking molecular na enzima XR da *Saccharomyces cerevisiae*

Um K_M mais elevado traduz uma menor afinidade, sendo que o contrário acontece com o score (“binding affinity” em Kcal/mol) resultante no Autodock Vina, quanto maior este parâmetro (em módulo) for melhor é a afinidade do substrato para enzima. Porém, provavelmente seria um erro comparar os scores resultantes de substratos diferentes na mesma enzima. Isto porque este score do Vina traduz a soma de todas as variações de energia de Gibbs de gauss, hidrofobicidade, repulsão, pontes de hidrogénio e número de ângulos de



torção de rotação livre do ligando, obtendo a energia assim a energia total de ligação, a qual o algoritmo do Vina minimiza tentando obter o melhor score de ligação [101]. Assim, quanto maior for o substrato sujeito ao docking nessa enzima, mesmo não sendo um substrato catalisado pela enzima, provavelmente maior será o score resultante, o que é incorreto inferir que é um resultado melhor. No caso da XR da *Saccharomyces cerevisiae*, os substratos com o mesmo número de carbonos apresentam os mesmos scores aproximadamente, sendo que este score aumenta conforme o número de carbonos aumenta (Tabela 4).

Tabela 4 – Scores do *docking* de 5 diferentes açúcares com diferentes tamanhos na XR da *Saccharomyces cerevisiae*.

Substrato (açúcar)	Score do <i>docking</i>
D-gliceroaldeído	-3,9
L-gliceroaldeído	-3,8
D-xilose	-5,0
L-arabinose	-5,0
D-glucose	-5,4

A D-xilose e a L-arabinose, ambas pentoses, apresentam um K_M de 27,9mM e 32,63mM respectivamente, enquanto o D-gliceroaldeído e o L-gliceroaldeído, ambos constituídos por 3 carbonos, apresentam K_M bastante distintos dos das pentoses como é possível observar pelo gráfico da Figura 25, apesar do K_M do L-gliceroaldeído ser semelhante ao da glucose (Figura 25).

Pelo gráfico da Figura 25 também se observa que parece não existir uma relação linear entre o número de carbonos e o K_M , apenas que é possível que os substratos com o mesmo número de carbonos apresentem K_M relativamente parecidos (Figura 25). O ideal seria ter dados sobre o K_M de mais uma hexose para verificar se o seu K_M se assemelha ao da glucose, isto porque o K_M deste hidrocarboneto é semelhante com o do L-gliceroaldeído.

Número de carbonos dos substratos vs K_M

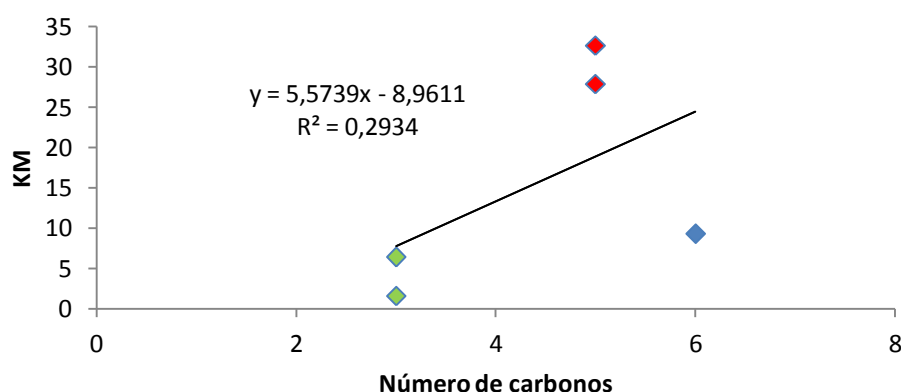


Figura 25 - Gráfico da relação entre o número de carbonos dos substratos da xilose redutase da *S.cerevisiae* e os seus respectivos K_M nesta enzima, com uma regressão linear $y = 5,5739x - 8,9611$ e um R² de 0,2934. A vermelho encontram-se as pentoses e a verde os substratos com 3 carbonos. O ponto azul é referente à glucose (6C).

3.10.1 Visualização dos resultados do docking da *Saccharomyces cerevisiae*

Apesar dos scores dos diferentes substratos nesta enzima não poderem ser comparados como já foi descrito, é possível visualizar e interpretar os resultados do docking e possíveis interações que estes açúcares fazem no centro ativo desta XR. Para todos os substratos analisados foi detetado que todos eles estabelecem ligações por pontes de hidrogénio com os resíduos His-111 e Asn-313. Porém, nas trioses, foi encontrada também uma interação destes açúcares com a cadeia lateral da tirosina (Tyr-49) (Figura 26). Na hexose, existem provavelmente pontes salinas com o resíduo Asp-50. Kavanagh KL et al [118] sugeriram que no centro catalítico da xilose redutase da *Candida tenuis*, os resíduos que estão envolvidos na ligação da xilose quando se encontra o NADP⁺ são o Tyr-51, Asp-50, Trp-23, His-113 e Asn-309. Isto sugere que os resíduos His-111 e Asn-313 que interagem com os substratos que foram sujeitos ao *docking*, devem ter um papel crucial na ligação de todos os açúcares e não apenas na xilose. É possível observar que as trioses estabelecem um maior número de ligações por pontes de hidrogénio com os resíduos envolventes e que por sua vez são aqueles que possuem um K_M mais baixo.

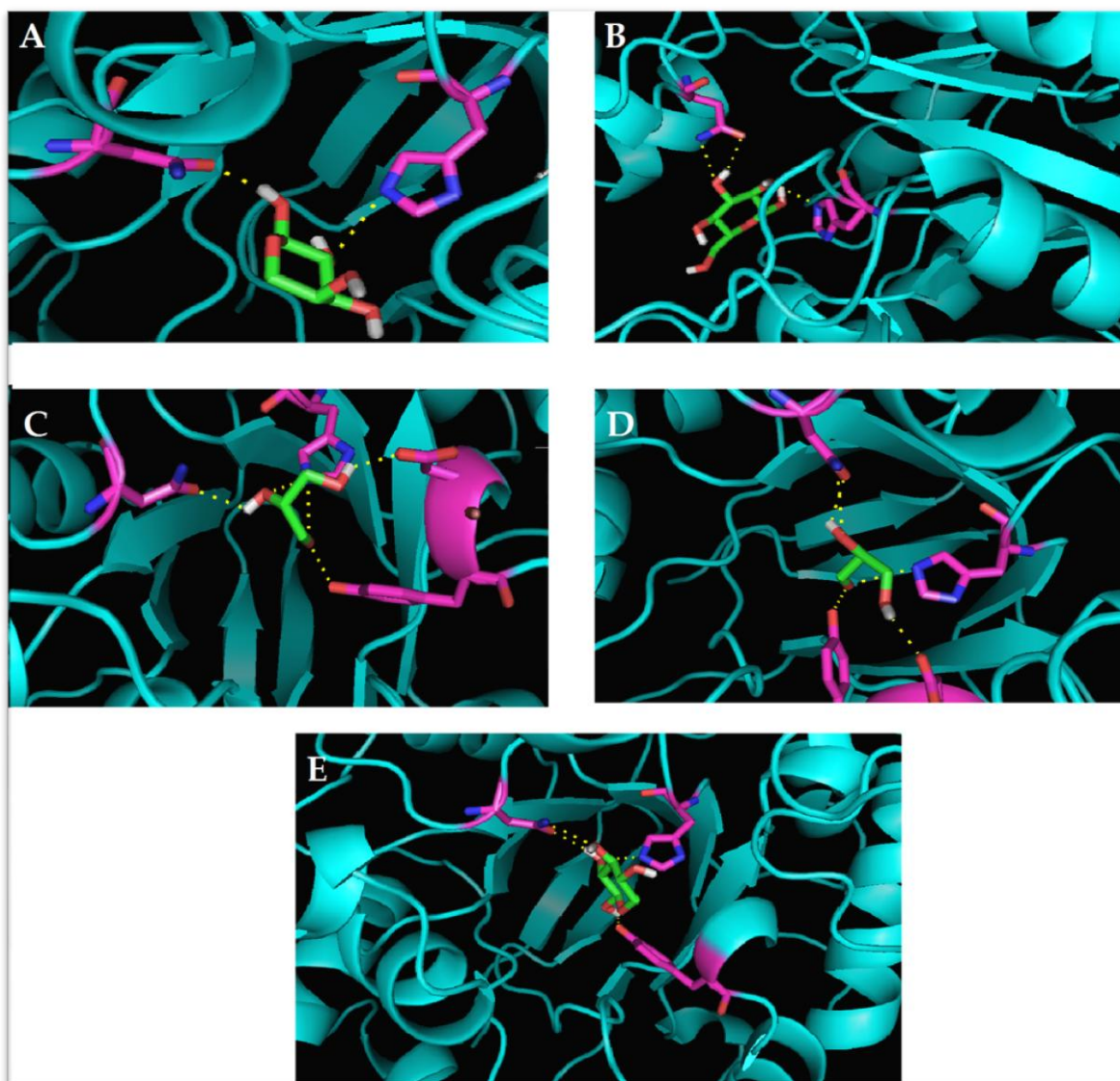


Figura 26 - Docking molecular da L-arabinose (A), D-glucose (B), D-gliceraldeído (C), L-gliceraldeído (D) e da D-xilose (E) no centro ativo da xilose redutase do organismo *S.cerevisiae*. O tracejado amarelo representa os contactos polares com do substrato com as cadeias laterais dos aminoácidos da proteína.

3.11 Docking do xilitol nas xilitol desidrogenases: propriedades, análise visual e química

Pela análise do score e visual de todas as xilitol desidrogenases sujeitas ao *docking* molecular, foi a XDH do organismo *Rhizomucor pusillus* que apresenta um maior score: 5,1 (tabela 5). Isto pode ser um indicador de que esta enzima estabelece ligações mais fortes e têm uma maior afinidade para o xilitol (Figura 27).



Figura 27 Centro ativo da xilitol desidrogenase do organismo *Rhizomucor pusillus* com o xilitol depois de ter sido sujeito ao processo de *docking*. É possível visualizar que os resíduos 300R, 156E e 52Y estabelecem ligações por pontes de hidrogénio (representadas a tracejado amarelo) com o xilitol.

Os 13 mutantes da *Scheffersomyces stipitis* apresentam scores distintos. As mutações D207A/I208R, D207AR/I208R/F209S e F209S são as que apresentam melhor score:4,6. Como o score da enzima não mutada foi de 4,6, o mais elevado a par de algumas enzimas mutadas referidas anteriormente, evidência que algumas das mutações, apesar de resolverem o desequilíbrio metabólico causado por a XR e XDH não terem o mesmo cofator preferencial, podem afetar a ligação do xilitol, diminuindo assim a afinidade para este álcool. Os K_M também foram comparados com os scores, não apresentando nenhuma relação significativa, provavelmente relacionado com os mesmos motivos acima referentes à XR.



Tabela 5 – Tabela com todas as xilitol desidrogenases e os seus respetivos scores de docking do xilitol juntamente com as interações polares que estabelecem no centro ativo. A abreviatura “pt” é referente a ligação por ponte de hidrogénio.

Organismo	Score do Docking	Aminoácidos que estabelecem interações polares no centro ativo
<i>Blastobotrys</i>	-4,3	H(2 pt) E(1 pt)
<i>Candiada spa</i>	-4,2	R(3 pt) H(1 pt) S(1 pt)
<i>Candida spHA E145C</i>	-4,8	R(4 pt) H(1 pt)
<i>Candida tropicalis</i>	-4,4	R(3 pt) T(1 pt) Y(1 pt) S(1 pt)
<i>Rhizobium etli</i>	-4,4	R(1 pt) S(1 pt)
<i>Rhizomucor pusillus</i>	-5,1	R(1 pt) E(1 pt) Y(1 pt)
<i>Rhizomucor pusillus mutante D205A</i>	-4,1	R(1 pt) H(1 pt)
<i>Rhizomucor pusillus mutante D206A/ I206R</i>	-4,4	H(1 pt) E(1 pt) G(1 pt)
<i>Saccaromices cerevisiae</i>	-4,4	R(1 pt)
<i>Scheffersomyces stipitis</i>	-4,6	R(3 pt) S(2 pt) T(1 pt)
<i>Scheffersomyces stipitis D207A</i>	-4,2	R(1 pt) A(1pt) S(1 pt)
<i>Scheffersomyces stipitis D207A /F209S</i>	-4,2	R(1 pt) T(2 pt)
<i>Scheffersomyces stipitis D207A /I208R</i>	-4,6	R(2 pt) S(1 pt)
<i>Scheffersomyces stipitis D207A/I208R/F209S</i>	-4,6	R(3 pt) T(2 pt)
<i>Scheffersomyces stipitis D207A/I208R/F209S/N211R</i>	-4,1	R(3 pt) T(1 pt)
<i>Scheffersomyces stipitis D207A/I208R/F209S/N211R/S96C/S99C/Y102C</i>	-4,5	R(2 pt) H(1 pt) S(1 pt)
<i>Scheffersomyces stipitis D207A/I208R/F209T</i>	-4,2	H(1 pt) S(1 pt)
<i>Scheffersomyces stipitis D207A/I208R/F209Y</i>	-4,4	S(1 pt) T(1 pt)



<i>Scheffersomyces stipitis</i> F209S	-4,6	R(2 pt) A(1pt) S(1 pt) T(2 pt)
<i>Scheffersomyces stipitis</i> I208R	-4,2	S(1 pt) T(1 pt)
<i>Scheffersomyces stipitis</i> I208R/F209S	-4,4	H(1 pt) E(1 pt)
<i>Scheffersomyces stipitis</i> N211R	-4,2	H(1 pt) T(1 pt)
<i>Scheffersomyces stipitis</i> S96C/S99C/Y102C	-4,2	R(2 pt) A(1 pt)
<i>Serratia marcescens</i>	-4,9	K(1 pt) T(1 pt) G(1pt)



4. CONCLUSÃO E PERSPETIVAS FUTURAS

Os volumes das diferentes xilose redutases conseguiram ser bem calculados pelo *trj_cavity*, porém não demonstraram uma forte correlação com os K_M da D-xilose, apenas uma tendência para serem mais elevados quanto menor for o K_M .

Os resultados relativos ao *docking* molecular demonstraram que as xilose redutases dos organismos *Candida boidiini* e *Candida tenuis* e a xilitol desidrogenase do organismo *Rhizomucor pusillus*, foram aquelas que obtiveram um melhor score de *docking*, podendo estas enzimas servir de ponto de partida para eventuais estudos estruturais que visem melhorar o processo de fermentação da xilose. É também possível concluir que a função score do Autodock Vina não é viável para comparar *binding affinitys* entre vários substratos da mesma enzima, mas que, apesar desta incapacidade, este programa demonstrou ser rápido e com boas previsões de conformação dos ligandos no centro ativo. Os resíduos His-111 e Asn-320 demonstram ter um papel crucial na reação de redução da xilose em xilitol e também em todos os outros açúcares reduzidos nesta enzima.

Em conclusão, a maioria dos trabalhos realizados para melhorar o processo de fermentação têm-se baseado numa abordagem metabólica, em trocas de cofator ou até mesmo a nível estrutural mas redirecionando-se novamente para o NADH e o NADPH. Como foi aqui demonstrado, as trocas de cofator preferencial podem afetar a afinidade do açúcar ou do xilitol para a enzima. Estas alterações melhoram significativamente a produção de etanol porém se conseguissem ser conjugadas com melhoramentos a afinidades do substrato provavelmente os resultados da obtenção de etanol seriam ainda mais promissores. Assim, este trabalho pode servir de ponto de partida para eventuais estudos estruturais, quer aperfeiçoamento dos aqui realizados quer novos, e experimentais relativos à D-xilose e ao xilitol, visando posteriormente conjugar com a parte metabólica deste problema.

Para um trabalho futuro, o uso de dinâmicas moleculares para medir o volume em diferentes estados da catálise enzima pode ser útil para a obtenção de resultados mais precisos e, posteriormente, se é compensatório aumentar o volume experimentalmente para conferir à D-xilose uma maior afinidade na XR.



REFERÊNCIAS BIBLIOGRÁFICAS

- 1- Khandekar, M.L., Murty, T.S., & Chittibabu, P. (2005). The global warming debate: a review of the state of science. *Pure Appl Geophys*, 162, 1557–1586.
- 2- Galbe, M., & Zacchi, G. (2002). A review of the production of ethanol from softwood. *Appl Microbiol Biotechnol*, 59, 618–628.
- 3- Moysés, D. N., Reis, V. C. B., de Almeida, J. R. M., de Moraes, L. M. P. & Torres, F. A. G. (2016). Xylose Fermentation by *Saccharomyces cerevisiae*: Challenges and Prospects. *Int J Mol Sci*, 17, 207.
- 4- Van Maris, A.J.A., Abbott, D.A., Bellissimi, E. Van den Brink, J., Kuyper, M., Luttik, M.A., Wisselink, H.W., Scheffers, W.A., van Dijken, J.P., & Pronk, J.T. (2006). Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: Current status. *Antonie Van Leeuwenhoek*, 90, 391–418.
- 5- What is Ethanol? - Formula, Structure & Uses. Disponível em: <https://study.com/academy/lesson/what-is-ethanol-formula-structure-uses.html>,
Acedido em: 2018-01-22
- 6- Olsson, L., & Hahn-Hägerdal, B. (1996). Fermentation of lignocellulosic hydrolyates for ethanol production. *Enzym. Microb. Technol*, 18, 312–331.
- 7- Langan, P. *et al.* (2014). L-Arabinose Binding, Isomerization, and Epimerization by D-Xylose Isomerase: X-Ray/Neutron Crystallographic and Molecular Simulation Study. *Structure*, 22, 1287–1300.
- 8- Duina, A. A., Miller, M. E., & Keeney, J. B. (2014). Budding yeast for budding geneticists: a primer on the *Saccharomyces cerevisiae* model system. *Genetics*, 197(1), 33-48.
- 9- Chiang, L.C., Gong C.S., Chen L.F., & Tsao G.T. (1981). D-Xylulose fermentation to ethanol by *Saccharomyces cerevisiae*. *Appl Environ Microbiol*, 42(2), 284–289.
- 10- Batt, C. A., Caryallo, S., Easson Jr, D. D., Akedo, M., & Sinskey, A. J. (1986). Direct evidence for a xylose metabolic pathway in *Saccharomyces cerevisiae*. *Biotechnology and bioengineering*, 28(4), 549-553.
- 11- Toivari, M. H., Salusjärvi, L., Ruohonen, L., & Penttilä, M. (2004). Endogenous xylose pathway in *Saccharomyces cerevisiae*. *Applied and environmental microbiology*, 70(6), 3681-3686.



- 12- Jin, Y. S., Ni, H., Laplaza, J. M., & Jeffries, T. W. (2003). Optimal growth and ethanol production from xylose by recombinant *Saccharomyces cerevisiae* require moderate D-xylokinase activity. *Applied and environmental microbiology*, 69(1), 495-503.
- 13- Kötter, P., & Ciriacy, M. (1993). Xylose fermentation by *Saccharomyces cerevisiae*. *Applied microbiology and biotechnology*, 38(6), 776-783.
- 14- Eliasson, A., Christensson, C., Wahlbom, C. F., & Hahn-Hägerdal, B. (2000). Anaerobic xylose fermentation by recombinant *Saccharomyces cerevisiae* carrying *XYL1*, *XYL2*, and *XKS1* in mineral medium chemostat cultures. *Applied and environmental microbiology*, 66(8), 3381-3386.
- 15- Hahn-Hägerdal, B., Jeppsson, H., Skoog, K., & Prior, B. A. (1994). Biochemistry and physiology of xylose fermentation by yeasts. *Enzyme and Microbial Technology*, 16(11), 933-943.
- 16- Walfridsson, M., Anderlund, M., Bao, X., & Hahn-Hägerdal, B. (1997). Expression of different levels of enzymes from the *Pichia stipitis* *XYL1* and *XYL2* genes in *Saccharomyces cerevisiae* and its effects on product formation during xylose utilisation. *Applied microbiology and biotechnology*, 48(2), 218-224.
- 17- Gong, C. S., Chen, L. F., Flickinger, M. C., Chiang, L. C., & Tsao, G. T. (1981). Production of ethanol from D-xylose by using D-xylose isomerase and yeasts. *Applied and environmental microbiology*, 41(2), 430-436.
- 18- Kruckeberg, A. L. (1996). The hexose transporter family of *Saccharomyces cerevisiae*. *Archives of Microbiology*, 166(5), 283-292.
- 19- Sedlak, M., & Ho, N. W. (2004). Characterization of the effectiveness of hexose transporters for transporting xylose during glucose and xylose co- fermentation by a recombinant *Saccharomyces* yeast. *Yeast*, 21(8), 671-684.
- 20- Goncalves, D. L., Matsushika, A., Belisa, B., Goshima, T., Bon, E. P., & Stambuk, B. U. (2014). Xylose and xylose/glucose co-fermentation by recombinant *Saccharomyces cerevisiae* strains expressing individual hexose transporters. *Enzyme and microbial technology*, 63, 13-20.
- 21- Nijland, J. G., Shin, H. Y., Boender, L. G., de Waal, P. P., Klaassen, P., & Driessen, A. J. (2017). Improved xylose metabolism by a *CYC8* mutant of *Saccharomyces cerevisiae*. *Applied and environmental microbiology*, AEM-00095.
- 22- Meinander, N. Q., Boels, I., & Hahn-Hägerdal, B. (1999). Fermentation of xylose/glucose mixtures by metabolically engineered *Saccharomyces cerevisiae*



- strains expressing XYL1 and XYL2 from *Pichia stipitis* with and without overexpression of TAL1. *Bioresource technology*, 68(1), 79-87.
- 23- Matsushika, A., Watanabe, S., Kodaki, T., Makino, K., & Sawayama, S. (2008). Bioethanol production from xylose by recombinant *Saccharomyces cerevisiae* expressing xylose reductase, NADP+-dependent xylitol dehydrogenase, and xylulokinase. *Journal of bioscience and bioengineering*, 105(3), 296-299.
- 24- Watanabe, S., Kodaki, T., & Makino, K. (2005). Complete reversal of coenzyme specificity of xylitol dehydrogenase and increase of thermostability by the introduction of structural zinc. *Journal of Biological Chemistry*, 280(11), 10340-10349.
- 25- Watanabe, S., Pack, S. P., Saleh, A. A., Annaluru, N., Kodaki, T., & Makino, K. (2007). The positive effect of the decreased NADPH-preferring activity of xylose reductase from *Pichia stipitis* on ethanol production using xylose-fermenting recombinant *Saccharomyces cerevisiae*. *Bioscience, biotechnology, and biochemistry*, 71(5), 1365-1369.
- 26- Almeida, J. R., Bertilsson, M., Hahn-Hägerdal, B., Lidén, G., & Gorwa-Grauslund, M. F. (2009). Carbon fluxes of xylose-consuming *Saccharomyces cerevisiae* strains are affected differently by NADH and NADPH usage in HMF reduction. *Applied microbiology and biotechnology*, 84(4), 751-761.
- 27- Sonderegger, M., Schümperli, M., & Sauer, U. (2004). Metabolic engineering of a phosphoketolase pathway for pentose catabolism in *Saccharomyces cerevisiae*. *Applied and environmental microbiology*, 70(5), 2892-2897.
- 28- Roca, C., Nielsen, J., & Olsson, L. (2003). Metabolic engineering of ammonium assimilation in xylose-fermenting *Saccharomyces cerevisiae* improves ethanol production. *Applied and environmental microbiology*, 69(8), 4732-4736.
- 29- Wahlbom, C. F., & Hahn-Hägerdal, B. (2002). Furfural, 5- hydroxymethyl furfural, and acetoin act as external electron acceptors during anaerobic fermentation of xylose in recombinant *Saccharomyces cerevisiae*. *Biotechnology and bioengineering*, 78(2), 172-178.
- 30- Jeppsson, M., Johansson, B., Hahn-Hägerdal, B., & Gorwa-Grauslund, M. F. (2002). Reduced oxidative pentose phosphate pathway flux in recombinant xylose-utilizing *Saccharomyces cerevisiae* strains improves the ethanol yield from xylose. *Applied and environmental microbiology*, 68(4), 1604-1609.



- 31- Khattab, S. M. R., Watanabe, S., Saimura, M., Afifi, M. M., Zohri, A. N. A., Abdul-Raouf, U. M., & Kodaki, T. (2011). Construction of a novel strictly NADPH-dependent *Pichia stipitis* xylose reductase by site-directed mutagenesis for effective bioethanol production. In *Zero-Carbon Energy Kyoto 2010* (pp. 117-122). Springer, Tokyo.
- 32- Khattab, S. M. R., Watanabe, S., Saimura, M., & Kodaki, T. (2011). A novel strictly NADPH-dependent *Pichia stipitis* xylose reductase constructed by site-directed mutagenesis. *Biochemical and biophysical research communications*, 404(2), 634-637.
- 33- Khattab, S. M. R., Saimura, M., & Kodaki, T. (2013). Boost in bioethanol production using recombinant *Saccharomyces cerevisiae* with mutated strictly NADPH-dependent xylose reductase and NADP⁺-dependent xylitol dehydrogenase. *Journal of biotechnology*, 165(3-4), 153-156.
- 34- Cavasotto, C. N., & Phatak, S. S. (2009). Homology modeling in drug discovery: current trends and applications. *Drug discovery today*, 14(13-14), 676-683.
- 35- Parida, B. K., Panda, P. K., Misra, N., & Mishra, B. K. (2015). MaxMod: a hidden Markov model based novel interface to MODELLER for improved prediction of protein 3D models. *Journal of molecular modeling*, 21(2), 30.
- 36- UniProtKB/Swiss-Prot UniProt release 2019_01. Disponível em: <https://www.uniprot.org/statistics/Swiss-Prot>., Acedido em: 2019-01-18.
- 37- Chou, K. C. (2004). Structural bioinformatics and its impact to biomedical science. *Current medicinal chemistry*, 11(16), 2105-2134.
- 38- Bioinformatics Resource Portal. Disponível em: <https://web.expasy.org/docs/relnotes/relstat.html>., Acedido em: 2019-01-18
- 39- PDB Statistics: Overall Growth of Released Structures Per Year. Disponível em: <https://www.rcsb.org/stats/growth/overall>., Acedido em 2019-01-18.
- 40- Fiser, A., & Do, R. K. G. (2000). Modeling of loops in protein structures. *Protein science*, 9(9), 1753-1773.
- 41- Chou, K. C., & Cai, Y. D. (2002). Using functional domain composition and support vector machines for prediction of protein subcellular location. *Journal of biological chemistry*, 277(48), 45765-45769.
- 42- Fiser, A., & Šali, A. (2003). Modeller: generation and refinement of homology-based protein structure models. In *Methods in enzymology* (Vol. 374, pp. 461-491). Academic Press.



- 43- Vyas, V. K., Ukawala, R. D., Ghate, M., & Chintla, C. (2012). Homology modeling a fast tool for drug discovery: current perspectives. *Indian journal of pharmaceutical sciences*, 74(1), 1.
- 44- Eswar, N., Webb, B., Marti- Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., ... & Sali, A. (2006). Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics*, 15(1), 5-6.
- 45- Sánchez, R., & Šali, A. (1999). Comparative protein structure modeling in genomics. *Journal of Computational Physics*, 151(1), 388-401.
- 46- Söding, J. (2004). Protein homology detection by HMM–HMM comparison. *Bioinformatics*, 21(7), 951-960.
- 47- Wang, Q., Canutescu, A. A., & Dunbrack Jr, R. L. (2008). SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling. *Nature protocols*, 3(12), 1832.
- 48- Song, Y., DiMaio, F., Wang, R. Y. R., Kim, D., Miles, C., Brunette, T. J., ... & Baker, D. (2013). High-resolution comparative modeling with RosettaCM. *Structure*, 21(10), 1735-1742.
- 49- Krivov, G. G., Shapovalov, M. V., & Dunbrack, R. L. (2009). Improved prediction of protein side- chain conformations with SCWRL4. *Proteins: Structure, Function, and Bioinformatics*, 77(4), 778-795.
- 50- Webb, B., & Sali, A. (2014). Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics*, 47(1), 5-6.
- 51- Fischer, E. (1894). Effects of configuration on enzyme activity. *Ber Dtsch Chem Ges*, 27, 2985-2993
- 52- Paramo, T., East, A., Garzón, D., Ulmschneider, M. B., & Bond, P. J. (2014). Efficient characterization of protein cavities within molecular simulation trajectories: trj_cavity. *Journal of chemical theory and computation*, 10(5), 2151-2164.
- 53- Henrich, S., Salo- Ahen, O. M., Huang, B., Rippmann, F. F., Cruciani, G., & Wade, R. C. (2010). Computational approaches to identifying and characterizing protein binding sites for ligand design. *Journal of Molecular Recognition: An Interdisciplinary Journal*, 23(2), 209-219.
- 54- Gutteridge, A., & Thornton, J. M. (2005). Understanding nature's catalytic toolkit. *Trends in biochemical sciences*, 30(11), 622-629.



- 55-Laskowski, R. A., Luscombe, N. M., Swindells, M. B., & Thornton, J. M. (1996). Protein clefts in molecular recognition and function. *Protein science: a publication of the Protein Society*, 5(12), 2438.
- 56-Bartlett, G. J., Porter, C. T., Borkakoti, N., & Thornton, J. M. (2002). Analysis of catalytic residues in enzyme active sites. *Journal of molecular biology*, 324(1), 105-121.
- 57-Harris, T. K., & Turner, G. J. (2002). Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB life*, 53(2), 85-98.
- 58-Levitt, D. G., & Banaszak, L. J. (1992). POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. *Journal of molecular graphics*, 10(4), 229-234.
- 59-Liang, J., Woodward, C., & Edelsbrunner, H. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein science*, 7(9), 1884-1897.
- 60-Brady, G. P., & Stouten, P. F. (2000). Fast prediction and visualization of protein binding pockets with PASS. *Journal of computer-aided molecular design*, 14(4), 383-401.
- 61-Venkatachalam, C. M., Jiang, X., Oldfield, T., & Waldman, M. (2003). LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *Journal of Molecular Graphics and Modelling*, 21(4), 289-307.
- 62-Simões, T., Lopes, D., Dias, S., Fernandes, F., Pereira, J., Jorge, J., ... & Gomes, A. (2017, December). Geometric detection algorithms for cavities on protein surfaces in molecular graphics: a survey. In *Computer Graphics Forum* (Vol. 36, No. 8, pp. 643-683).
- 63-Kuhn, D., Weskamp, N., Schmitt, S., Hüllermeier, E., & Klebe, G. (2006). From the similarity analysis of protein cavities to the functional classification of protein families using cavbase. *Journal of molecular biology*, 359(4), 1023-1044.
- 64-Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of medicinal chemistry*, 28(7), 849-857.
- 65-Laurie, A. T., & Jackson, R. M. (2005). Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites. *Bioinformatics*, 21(9), 1908-1916.



- 66- Harris, R., Olson, A. J., & Goodsell, D. S. (2008). Automated prediction of ligand binding sites in proteins. *Proteins: structure, function, and bioinformatics*, 70(4), 1506-1517.
- 67- Najmanovich, R., Kurbatova, N., & Thornton, J. (2008). Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics*, 24(16), i105-i111.
- 68- Schmitt, S., Kuhn, D., & Klebe, G. (2002). A new method to detect related function among proteins independent of sequence and fold homology. *Journal of molecular biology*, 323(2), 387-406.
- 69- Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., & Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, 19(1), 163-164.
- 70- Pupko, T., Bell, R. E., Mayrose, I., Glaser, F., & Ben-Tal, N. (2002). Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*, 18(suppl_1), S71-S77.
- 71- Brylinski, M., & Skolnick, J. (2008). A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proceedings of the National Academy of sciences*, 105(1), 129-134.
- 72- Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., ... & Biedermannova, L. (2012). CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS computational biology*, 8(10), e1002708.
- 73- Laurent, B., Chavent, M., Cragolini, T., Dahl, A. C. E., Pasquali, S., Derreumaux, P., ... & Baaden, M. (2014). Epock: rapid analysis of protein pocket dynamics. *Bioinformatics*, 31(9), 1478-1480.
- 74- Raunest, M., & Kandt, C. (2011). dxTuber: detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. *Journal of Molecular Graphics and Modelling*, 29(7), 895-905.
- 75- Schmidtke, P., Bidon-Chanal, A., Luque, F. J., & Barril, X. (2011). MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*, 27(23), 3276-3285.
- 76- Gong, S., Park, C., Choi, H., Ko, J., Jang, I., Lee, J., ... & Bhak, J. (2005). A protein domain interaction interface database: InterPare. *BMC bioinformatics*, 6(1), 207.



- 77- Kellogg, G. E., Semus, S. F., & Abraham, D. J. (1991). HINT: a new method of empirical hydrophobic field calculation for CoMFA. *Journal of computer-aided molecular design*, 5(6), 545-552.
- 78- Verdonk, M. L., Cole, J. C., & Taylor, R. (1999). SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *Journal of molecular biology*, 289(4), 1093-1108.
- 79- Heiden, W., Moeckel, G., & Brickmann, J. (1993). A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. *Journal of computer-aided molecular design*, 7(5), 503-514.
- 80- Nayal, M., & Honig, B. (2006). On the nature of cavities on protein surfaces: application to the identification of drug- binding sites. *Proteins: Structure, Function, and Bioinformatics*, 63(4), 892-906.
- 81- Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry: AABC*, 8, 37.
- 82- Jain, A. N., & Nicholls, A. (2008). Recommendations for evaluation of computational methods. *Journal of computer-aided molecular design*, 22(3-4), 133-139.~
- 83- Lengauer, T., & Rarey, M. (1996). Computational methods for biomolecular docking. *Current opinion in structural biology*, 6(3), 402-406.
- 84- Alonso, H., Bliznyuk, A. A., & Gready, J. E. (2006). Combining docking and molecular dynamic simulations in drug design. *Medicinal research reviews*, 26(5), 531-568.
- 85- Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery*, 3(11), 935.
- 86- Levinthal, C., Wodak, S. J., Kahn, P., & Dadvanian, A. K. (1975). Hemoglobin interaction in sickle cell fibers. I: Theoretical approaches to the molecular contacts. *Proceedings of the National Academy of Sciences*, 72(4), 1330-1334.
- 87- Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: a review. *Biophysical reviews*, 9(2), 91-102.
- 88- Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*, 19(14), 1639-1662.



- 89- Ewing, T. J., Makino, S., Skillman, A. G., & Kuntz, I. D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *Journal of computer-aided molecular design*, 15(5), 411-428.
- 90- Rarey, M., Kramer, B., Lengauer, T., & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *Journal of molecular biology*, 261(3), 470-489.
- 91- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., ... & Shaw, D. E. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of medicinal chemistry*, 47(7), 1739-1749.
- 92- Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein–ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics*, 52(4), 609-623.
- 93- Abagyan, R., Totrov, M., & Kuznetsov, D. (1994). ICM—a new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *Journal of computational chemistry*, 15(5), 488-506.
- 94- McMartin, C., & Bohacek, R. S. (1997). QXP: powerful, rapid computer algorithms for structure-based drug design. *Journal of computer-aided molecular design*, 11(4), 333-344.
- 95- Jain, A. N. (2003). Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *Journal of medicinal chemistry*, 46(4), 499-511.
- 96- Isabella, A. Guedes, Camila S. de Magalhães, & Laurent E. Dardenn. (2014). Receptor–ligand molecular docking, *Biophys Rev*, 6(1), 75–87.
- 97- Handoko, S. D., Ouyang, X., Su, C. T. T., Kwoh, C. K., & Ong, Y. S. (2012). QuickVina: accelerating AutoDock Vina using gradient-based heuristics for global optimization. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 9(5), 1266-1272.
- 98- Novič, M., Tibaut, T., Anderluh, M., Borišek, J., & Tomašič, T. (2016). The comparison of docking search algorithms and scoring functions: an overview and case studies. In *Methods and Algorithms for Molecular Docking-Based Drug Design and Discovery* (pp. 99-127). IGI Global.



- 99- Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., & Corbeil, A. C. (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British journal of pharmacology*, 153(S1), S7-S26.
- 100- Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Current computer-aided drug design*, 7(2), 146-157.
- 101- Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), 455-461.
- 102- Ng, M. C., Fong, S., & Siu, S. W. (2015). PSOVina: The hybrid particle swarm optimization algorithm for protein–ligand docking. *Journal of bioinformatics and computational biology*, 13(03), 1541007.
- 103- Koebel, M. R., Schmadeke, G., Posner, R. G., & Sirimulla, S. (2016). AutoDock VinaXB: implementation of XBSF, new empirical halogen bond scoring function, into AutoDock Vina. *Journal of cheminformatics*, 8(1), 27.
- 104- Hyndman, D., Bauman, D. R., Heredia, V. V., & Penning, T. M. (2003). The aldo-keto reductase superfamily homepage. *Chemico-biological interactions*, 143, 621-631.
- 105- Jez, J. M., & Penning, T. M. (2001). The aldo-keto reductase (AKR) superfamily: an update. *Chemico-biological interactions*, 130, 499-525.
- 106- BENNETT, M. J., SCHLEGEL, B. P., LEWIS, M., & Trevor, M. (1997). Comparative anatomy of the aldo–keto reductase superfamily. *Biochemical Journal*, 326(3), 625-636.
- 107- Jez, J. M., Flynn, T. G., & Penning, T. M. (1997). A new nomenclature for the aldo-keto reductase superfamily. *Biochemical pharmacology*, 54(6), 639-647.
- 108- Nordling, E., Jörnvall, H., & Persson, B. (2002). Medium- chain dehydrogenases/reductases (MDR) Family characterizations including genome comparisons and active site modelling. *European journal of biochemistry*, 269(17), 4267-4276.
- 109- Riveros Rosas, H., Julián Sánchez, A., Villalobos Molina, R., Pardo, J. P., & Piña, E. (2003). Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. *European journal of biochemistry*, 270(16), 3309-3334.



- 110- Wang, J. F., Wei, D. Q., Lin, Y., Wang, Y. H., Du, H. L., Li, Y. X., & Chou, K. C. (2007). Insights from modeling the 3D structure of NAD (P) H-dependent D-xylose reductase of *Pichia stipitis* and its binding interactions with NAD and NADP. *Biochemical and biophysical research communications*, 359(2), 323-329.
- 111- Kratzer, R., Wilson, D. K., & Nidetzky, B. (2006). Catalytic mechanism and substrate selectivity of aldo- keto reductases: Insights from structure- function studies of *Candida tenuis* xylose reductase. *IUBMB life*, 58(9), 499-507.
- 112- Kratzer, R., & Nidetzky, B. (2005). Electrostatic stabilization in a pre-organized polar active site: the catalytic role of Lys-80 in *Candida tenuis* xylose reductase (AKR2B5) probed by site-directed mutagenesis and functional complementation studies. *Biochemical journal*, 389(2), 507-515.
- 113- Brenda. Disponível em: <https://www.brenda-enzymes.org/>., Acedido em: 15/2/2018
- 114- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., ... & Martin, M. J. (2004). UniProt: the universal protein knowledgebase. *Nucleic acids research*, 32(suppl_1), D115-D119.
- 115- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Thompson, J. D. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), 539.
- 116- Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., ... & Bryant, S. H. (2009). The NCBI biosystems database. *Nucleic acids research*, 38(suppl_1), D492-D496
- 117- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E. (2000). The protein data bank. *Nucleic acids research*, 28(1), 235-242.
- 118- Kavanagh, K. L., Klimacek, M., Nidetzky, B., & Wilson, D. K. (2002). The structure of apo and holo forms of xylose reductase, a dimeric aldo-keto reductase from *Candida tenuis*. *Biochemistry*, 41(28), 8785-8795.
- 119- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16), 2785-2791.



- 120- O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of cheminformatics*, 3(1), 33.
- 121- The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.
- 122- Copeland, R. A. (2004). *Enzymes: a practical introduction to structure, mechanism, and data analysis*. John Wiley & Sons.
- 123- Coppes, Z., Martínez, G., & Hirschhorn, M. (1992). pH and Temperature effects on the Km values of muscle lactate dehydrogenase isozyme LDH-A4 from fishes of the family Sciaenidae (Perciformes). *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, 103(4), 869-874
- 124- Zeng, Q. K., Du, H. L., Wang, J. F., Wei, D. Q., Wang, X. N., Li, Y. X., & Lin, Y. (2009). Reversal of coenzyme specificity and improvement of catalytic efficiency of Pichiastipitis xylose reductase by rational site-directed mutagenesis. *Biotechnology letters*, 31(7), 1025-1029.
- 125- Liang, L., Zhang, J., & Lin, Z. (2007). Altering coenzyme specificity of Pichia stipitis xylose reductase by the semi-rational approach CASTing. *Microbial Cell Factories*, 6(1), 36.
- 126- Fernandes, S., Tuohy, M. G., & Murray, P. G. (2009). Xylose reductase from the thermophilic fungus Talaromyces emersonii: cloning and heterologous expression of the native gene (Texr) and a double mutant (Texr K271R+ N273D) with altered coenzyme specificity. *Journal of biosciences*, 34(6), 881-890
- Rose, A. S., & Hildebrand, P. W. (2015). NGL Viewer: a web application for molecular visualization. *Nucleic acids research*, 43(W1), W576-W579.
- 127- Rose, A. S., & Hildebrand, P. W. (2015). NGL Viewer: a web application for molecular visualization. *Nucleic acids research*, 43(W1), W576-W579.
- 128- Durrant, J. D., de Oliveira, C. A. F., & McCammon, J. A. (2011). POVME: an algorithm for measuring binding-pocket volumes. *Journal of Molecular Graphics and Modelling*, 29(5), 773-776.